The London School of Economics and Political Science

Essays on the Economics of Taxation

Michael Carlos Best

A thesis submitted to the Department of Economics of
the London School of Economics for the degree of
Doctor of Philosophy, London, May 2014

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 49,219 words.

Statement of conjoint work

I confirm that chapters 1 & 3 were jointly co-authored with Professor Henrik Kleven and I contributed 50% of this work.

Dedication

Completing this thesis would not have been possible without the support of my family, my friends, and my colleagues. Though there are more people who deserve thanks than I can name here, I would like to particularly thank my parents, Edward and Nubia, for their patient love and support through the many ups and downs of my PhD; my grandparents Geoffrey and Marigold who always showed interest in my work, and gave a refugee student shelter during his master's; my uncle Simon without whose generosity I would not be where I am today; Michelle, who always pushed me to strive for excellence, and refuse to accept failure; the usual suspects: Rob, Jeremy, Katharine, Emma, Coxy, Emma (2), Set, Ivan, Tom, Michelle (2), for always being there willing to hang out with a dishevelled student bum; and all the economics friends I've made along the way. Special thanks go to my supervisors Henrik, Oriana and Andrea, who somehow mistakenly saw some potential in me, nurtured it, and believed in it even when I was ready to give up.

London, 9 May 2014

Abstract

This thesis explores the way economic behaviour responds to taxation both theoretically and empirically.

Chapter 1 studies the impact of transation taxes on the housing market, using UK administrative data and quasi-experimental variation created by notches, tax reforms, and stimulus. Transaction taxes have large effects on house prices and purchases, and adjustments to tax changes are fast. A temporary elimination of transaction taxes stimulated housing market activity by 20% in the short run (timing and extensive responses) followed by a smaller slump in activity after the policy was withdrawn (timing response). The success of this stimulus program stems from the large distortions created by the tax in the first place.

Chapter 2 presents evidence on three ways in which firms affect workers' earnings responses in Pakistan. First, third-party reporting of salaries by employers reduces evasion. Second, firms' equilibrium salary-hours offers are tailored to aggregate worker preferences in response to adjustment costs in the labour market. Third, workers learn about the tax schedule from firms and become more responsive to taxation both contemporaneously (by 130%) and in subsequent years (by 100%). Third-party reporting does not eliminate misreporting: 19% of workers underreport their salaries, creating a loss of 5% of tax revenue, and indicating high returns to investments in improving enforcement.

Chapter 3 develops a theory of optimal income taxation allowing for career effects of current work effort on future wages. Such effects are empirically important, but have been ignored by the optimal tax literature. We provide analytical characterizations that depend on estimable entities, including the elasticity of future wages to current work effort. We explore the magnitude of this "career elasticity" in a meta-analysis of the empirical literature on the returns to work experience and tenure, and provide numerical simulations calibrated to US micro-data. Our results show that career effects have important qualitative and quantitative implications for optimal tax design.

Contents

List of Figures

List of Tables

CHAPTER 1

Housing Market Responses to Transaction Taxes: Evidence From Notches and Stimulus in the UK[1]

ABSTRACT. Using administrative data on all property transactions in the UK from 2004-2012, we provide evidence on the distortionary effects of property transaction taxes ("stamp duty") on the housing market. Two sources of quasi-experimental variation allow us to obtain compelling graphical results: (i) notches created by discontinuous jumps in tax liability at threshold property prices, (ii) time variation created by permanent reforms and temporary stimulus in specific price brackets. We present two broad findings. First, transaction taxes strongly affect the price, volume and timing of property transactions. Our findings imply that property transaction taxes are extremely distortionary, with Laffer rates as low as 4–7%. Second, temporary transaction tax cuts are an enormously effective form of fiscal stimulus. A temporary elimination of a 1% transaction tax increased housing market activity by 20% in the short run (due to both timing and extensive responses) and less than half of the stimulus effect was reversed after the tax was reintroduced (due to re-timing). Due to the complementarities between moving house and consumer spending, these stimulus effects translate into GDP effects of about 1 dollar per dollar of foregone revenue. This is considerably larger than what has been found for other forms of fiscal stimulus.

[1]We thank Tim Besley, Raj Chetty, Julie Cullen, Michael Devereux, Roger Gordon, Daniel Hamermesh, Benjamin Keys, Wojciech Kopczuk, Camille Landais, Attila Lindner, Bruce Meyer, Atif Mian, David Munroe, Emmanuel Saez, Andrei Shleifer, Joel Slemrod, and numerous seminar participants for very helpful comments and discussions. We would also like to thank the staff at Her Majesty's Revenue & Customs' (HMRC) datalab for access to the data and their support of this project. This work contains statistical data from HMRC which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information. All results have been screened by HMRC to ensure confidentiality is not breached. All remaining errors are the authors'.

1.1. Introduction

Taxes on asset transfers are widely discussed by economists and policy makers, but remain understudied (Campbell & Froot 1994; Poterba 2002; Matheson 2011; European Commission 2013). This debate has been particularly energetic in recent years as a result of the enormous turmoil in the financial and housing markets and their potential importance for the boom-bust cycle of the economy as a whole. In this paper we focus on the taxation of the transfer of residential property—a policy that is ubiquitous throughout the world, and raises substantial amounts of revenue in many countries.[2] We analyze the UK property transaction tax, known as the Stamp Duty Land Tax (SDLT), which is substantial both in terms of revenue and the distortions it introduces. Our analysis is based on unique access to administrative records covering the universe of property transactions combined with policy-induced quasi-experimental variation that allows us to obtain compelling and striking evidence on housing market responses to transaction taxes.

Our analysis delivers two broad findings. First, the transaction tax is highly distortionary across a range of margins, causing significant distortions to the price, volume and timing of property transactions. This finding raises significant questions about the suitability of a property transaction tax as a long run policy instrument. Second, temporarily eliminating the transaction tax as a stimulus policy during a recession is enormously effective. We find that such stimulus increases housing market activity dramatically, and that consumer expenditures complementary to moving house increase by roughly the amount of the tax cut. This finding is not specific to the elimination of a distortionary tax: reducing the cost of transacting a house in general appears to be a powerful stimulus policy. Beyond the UK, this finding lends support to the reasoning behind the homebuyer tax credit introduced by the 2009 Stimulus Bill in the US.

We exploit first-time access to administrative tax data on the universe of property transactions in the UK from 2004–2012, about 10 million property transactions, with rich tax return information on each transaction. Besides the quality of the data, two sources of

[2] As of 2012, 38 states in the US had a property transaction tax (Lincoln Institute of Land Policy, 2014). In 2010 in the OECD, Austria, Australia, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, Germany, Greece, Hungary, Ireland, Iceland, Israel, Italy, Japan, the Netherlands, Norway, Poland, Portugal, South Korea, Spain, Sweden, Turkey, the UK, and the USA imposed transaction taxes on property (Andrews *et al.*, 2011). Beyond the OECD Hong Kong, India, Pakistan, and Singapore also impose property transaction taxes.

quasi-experimental variation allow us to obtain compellingly identified and striking visual evidence on housing market responses to transaction taxes along a variety of margins. First, the UK stamp duty features large discontinuities in tax liability—*notches*—at cutoff property prices. For example, the tax rate jumps from 1% to 3% of the entire transaction price at a cutoff of £250,000 (about $400,000), creating an increase in tax liability of £5,000 (about $8,000) as the house price crosses this cutoff. Such notches create strong incentives for reducing house prices in a region above the cutoff to a point just below the cutoff, thereby creating a hole in the price distribution on the high-tax side and excess bunching in the price distribution on the low-tax side of the notch. This allows for non-parametric identification of house price responses to transaction taxes using a bunching approach (Saez 2010; Chetty *et al.* 2011; Kleven & Waseem 2013). Second, the UK stamp duty features subtantial time variation, including both permanent and temporary tax changes that affect specific price brackets but not others. For example, a *stamp duty holiday* lasting 16 months eliminated transaction taxes in a certain price range in order to provide stimulus to the housing market during the current recession. As we show, this provides an ideal setting for a difference-in-differences approach to evaluating both extensive responses (whether or not to buy a house) and timing responses (when to buy a house) to temporary stimulus.

Our empirical findings can be divided into four main categories. First, there is large and sharp bunching just below notch points combined with large holes above notch points in the distribution of house prices. Our bunching estimates imply that house prices respond by a factor of 2–5 times the size of the tax increase at the notch, with larger effects at the bottom than at the top of the price distribution. Since notches create extremely large implicit marginal tax rates in the vicinity of the cutoff, the large bunching responses are consistent with more modest elasticities of house prices with respect to marginal tax rate, around 0.1–0.3 across most notches. We show that these effects on the market value of transacted houses ("house prices") may be driven by both the demand for quality-adjusted units of housing and the price per unit (through price bargaining), but not by standard market-level price incidence and such effects are therefore not part of our estimates.[3]

[3]Our estimates of house price responses are analogous to the literature on taxable income responses (Saez *et al.* 2012), which combines real labor supply and wage bargaining effects (but does not include standard wage incidence).

Second, we consider the dynamics of house price responses using both anticipated and unanticipated changes in the location of notches. The dynamic adjustment of bunching and holes to changes in notches is very fast, with a new steady state emerging in about 3–4 months for unanticipated changes and almost immediately for anticipated changes. The remarkable sharpness of our dynamic findings suggests that agents in the housing market are less affected by optimization frictions (inattention, inertia, etc.) than for example agents in the labor market (Chetty *et al.* 2011; Chetty 2012; Kleven & Waseem 2013; Gelber *et al.* 2013).

Third, we find strong evidence of short-term timing responses to pre-announced tax changes that create *time notches* at cutoff dates. In the two weeks leading up to an anticipated tax increase, activity levels in the housing market increased by around 150%. Our sharp, non-parametric evidence on timing responses in the housing market contributes to previous findings that short-term timing responses may far exceed medium- or long-run responses (Auerbach 1988; Burman & Randolph 1994; Goolsbee 2000).

Fourth, we estimate medium-term timing and extensive margin responses using temporary and permanent tax reforms.[4] Temporary housing stimulus successfully boosts activity in the short run as transaction volumes in the treatment group clearly diverge from transaction volumes in a control group during the 16-month stamp duty holiday. A 1%-point cut in transaction taxes increases market activity by about 20% during the holiday. This effect combines a *timing effect* (intertemporal substitution by those who would have purchased a house anyway) and an *extensive margin effect* (house purchases that would not have taken place absent the tax holiday). We can separate the two effects by comparing treatments and controls following the removal of the stimulus policy. Consistent with a timing effect, activity levels in the treatment group drop by about 8% compared to the control group in the first year after the holiday, with no further reversal in the second year after the holiday. The total reversal effect due to re-timing is less than half of the total stimulus boost, in contrast to Mian & Sufi (2012) who find complete reversal within one year of a US stimulus program. Our estimates imply extremely large elasticities with respect to

[4]These are extensive responses for house *purchases* as opposed to house *ownership*. Hence, our estimates of extensive responses do not just capture movements between renting and owning, but also that existing homeowners make additional house purchases (and therefore move more) over their lifetime.

transaction tax rates, a medium-term elasticity (including intertemporal substitution) of more than 20 and a long-run extensive margin elasticity of more than 14.

Our results have implications for the design of fiscal stimulus. Even though higher transaction levels in the housing market (for a given aggregate housing stock) do not add *mechanically* to real economic activity, house purchases have important real effects. Besides the implications of homeowner mobility for housing and labor markets, moving house is associated with substantial household spending on repairs, renovations, durable goods (domestic appliances, consumer electronics, furnishing, etc.), and commissions to agents and lawyers. Using UK consumption survey data, we estimate conservatively that a house transaction triggers extra spending of about 5% of the house price. Combined with our estimated increase in transaction volume (20%) and the size of the tax cut (1% of the house price), this implies that the amount of extra economic activity per dollar of tax cut is about 1. This captures only the immediate stimulus effect of larger spending; it does not include potential multiplier effects or indirect effects of mobility. Compared to a large body of evidence on consumer responses to other forms of fiscal stimulus such as tax rebates (e.g. Shapiro & Slemrod 2003a,b; Johnson *et al.* 2006; Agarwal *et al.* 2007; Kreiner *et al.* 2012), our findings suggest that the spending impact of the UK housing stimulus program has been considerably larger. The large effect is due to the strong responsiveness of house purchases to transaction taxes along with the complementarities between moving house and consumer spending. More generally, reducing transactions costs in the housing market (using tax cuts or subsidies) may be a powerful form of stimulus

Our results also have implications for the long-run revenue and welfare effects of property transaction taxes. We estimate that the marginal cost of public funds for the stamp duty varies from 1.17 in the lowest tax bracket to 18 in the highest tax bracket. For comparison, this is orders of magnitude larger than for the personal income tax, where it varies from 1.1 at the bottom of the income distribution to 2.68 at the top of the distribution (Kleven & Kreiner 2006), and for the UK tax system as a whole where it is only about 1.2 (Dahlby 2008). Furthermore, our estimates imply that the revenue-maximizing tax rate—the Laffer rate—is only around 4–7%, and so transaction taxes above this modest level reduce revenue and are Pareto inefficient. It is important to note that while the presence of notches in the UK stamp duty makes it particularly distortionary, it is not the presence of notches *per se* that makes the transaction tax so inefficient. Rather, this is due to the underlying responsiveness to this type of tax and applies to smooth schedules as well.

Transaction taxes are understudied in the enormous empirical literature on taxation and our paper takes a step towards closing this gap. A small body of prior work has studied the effects of property transaction taxes on house prices and homeowner mobility in different countries (Benjamin *et al.* 1993; van Ommeren & van Leuvensteijn 2005; Besley *et al.* 2011; Dachis *et al.* 2012). Moreover, related to our first result (static house price bunching), two contemporaneous papers by Slemrod *et al.* (2012) and Kopczuk & Munroe (2013) find similar house price responses using US tax notches. A larger empirical literature has examined the impact of capital gains taxes on asset prices and asset transactions (e.g. Feldstein *et al.* 1980; Auerbach 1988; Burman & Randolph 1994) and some of this work has focused specifically on the taxation of housing capital gains (Cunningham & Engelhardt 2008; Shan 2011). Capital gains taxes and transaction taxes share the feature that tax liability is triggered by a transaction, with the key difference being that transaction taxes fall on the entire value of the asset and not just on the appreciation of the asset. In contrast to the rest of the literature, we are able to simultaneously exploit a large dataset of administrative tax records along with multiple sources of quasi-experimental variation from notches, tax reforms and stimulus. This provides us with compelling non-parametric identification of a broader set of responses (prices, timing, extensive margin) viewed both statically and dynamically. Exploiting these we are able to provide a more complete picture of the (large) distortions introduced by transaction taxes and to provide compelling evidence on the efficacy of policies that reduce transaction costs in the housing market as stimulus policy during recessions.

The paper proceeds as follows. Section 2 presents our conceptual framework, section 3 describes the context and data, section 4 estimates house price responses using notches, section 5 estimates timing and extensive responses using stimulus and permanent reforms, and section 6 concludes.

1.2. Conceptual Framework

To guide the empirical analysis, this section first develops a simple static model of a competitive housing market and then considers a dynamic extension of that model. The framework is deliberately unrealistic in some dimensions as our main goal is to build the most parsinomious model possible that is still general enough to demonstrate the key empirical effects. Appendix 1.7.1 extends our conceptual analysis to a setting with matching frictions and price bargaining.

1.2.1. A Static Model of the Housing Market. Agents choose whether or not to become homeowners (extensive margin) and how much housing to buy conditional on owning (intensive margin). Letting c denote units of a numeraire consumption good and h denote units of quality-adjusted housing stock, we consider the following parametrization of preferences

$$u\left(c,h\right)=c+\frac{A}{1+1/\varepsilon}\left(\frac{h}{A}\right)^{1+1/\varepsilon}-q\cdot I\left\{h>0\right\}, \tag{1.2.1}$$

where A,ε are parameters characterizing housing preferences and q is a fixed cost of entering the owner-occupied market including both transaction costs (search costs, broker fees, etc.) and the utility from renting instead of owning. We allow for heterogeneity in all of these parameters captured by a smooth density distribution $f\left(A,\varepsilon,q\right)$. The quasi-linear utility function conveniently eliminates income effects on housing demand as we will focus purely on the price effect.

As a baseline, consider a flat transaction tax rate t on the value of housing purchased. Denoting the price per unit of housing by p and income by y, the budget constraint is given by

$$c+\left(1+t\right)ph=y. \tag{1.2.2}$$

Conditional on owning ($h>0$), maximizing utility (1.2.1) with respect to the budget constraint (1.2.2) yields the following housing demand function

$$h^{*}=A\left(\left(1+t\right)p\right)^{\varepsilon}, \tag{1.2.3}$$

where ε is the price elasticity of housing demand. Indirect utility conditional on $h>0$ and exclusive of the fixed cost q can be defined as $v\left(\left(1+t\right)p,y\right)\equiv u\left(c^{*},h^{*}\right)+q$, while indirect utility conditional on $h=0$ is given by $u\left(y,0\right)=y$. The agent then enters the owner-occupied housing market iff

$$q \leq v\left(\left(1+t\right)p, y\right) - y \equiv q^* \tag{1.2.4}$$

Total housing demand is then given by

$$D\left(\left(1+t\right)p\right) = \int_A \int_\varepsilon \int_0^{q^*} h^* f\left(A, \varepsilon, q\right) dq d\varepsilon dA. \tag{1.2.5}$$

We will be agnostic about the details of the supply side and denote housing supply by $S\left(p\right)$. The equilibrium condition $D\left(\left(1+t\right)p\right) = S\left(p\right)$ determines the equilibrium price p as a function of $1+t$.

Now consider the introduction of a discrete jump Δt in the average transaction tax rate—a notch—at a cutoff property value. Denoting property value by $h_v \equiv ph$, the notched tax schedule can be written as $T\left(h_v\right) = t \cdot h_v + \Delta t \cdot h_v \cdot I\left\{h_v > \overline{h}_v\right\}$ where $\overline{h}_v$ is the cutoff and $I\left\{\cdot\right\}$ is an indicator for being above the cutoff. Figure 1 illustrates the implications of this notch in a budget set diagram (Panel A) and density distribution diagrams (Panels B-D). The budget set diagram (depicted in (h_v, c)-space) illustrates intensive responses among individuals with heterogeneous housing preferences A, but a specific demand elasticity ε. The notch creates bunching at the cutoff $\overline{h}_v$ by all individuals in a preference range $\left(\overline{A}, \overline{A} + \Delta\overline{A}\right)$, who would have bought houses on the segment $\left(\overline{h}_v, \overline{h}_v + \Delta\overline{h}_v\right)$ in the absence of the notch. The marginal bunching individual at $\overline{A} + \Delta\overline{A}$ is indifferent between the notch point $\overline{h}_v$ and the best interior location $\bar{h}_v^I$. No individual is willing to locate between $\overline{h}_v$ and $\bar{h}_v^I$, and hence this range is completely empty. The density distribution of property values corresponding to the budget set diagram (all A, one specific ε) is shown in Panel B. Since the behavioral response in Panels A-B depends on the size of the demand elasticity ε (and converges to zero for $\varepsilon = 0$), the density distribution in the full population (all A, ε) can be illustrated as in Panel C where some individuals are willing to buy just above the notch point.[5]

[5]Notice that the above characterization is based on a given price p per unit of housing. The tax-induced change in aggregate housing demand (from bunching as well as interior responses further up) will affect the equilibrium price, which by itself will shift indifference curves in Panel A (as they are depicted in (h_v, c)-space) and hence shift the density distribution of property values. The qualitative characterization above holds for any arbitrary price and therefore also for the new equilibrium price. The key insight is that, in this competitive model, price incidence occurs at the *market level* and therefore does not contribute to bunching and holes locally around notches. Appendix 1.7.1 considers a bargaining model where price incidence occurs at the *match level* in which case price incidence does create bunching and holes.

In addition to intensive responses, the notch creates extensive responses above the cutoff by individuals close to being indifferent between buying and not buying (with $q \approx q^*$). However, such extensive responses will be negligible *just* above the cutoff. This can be seen by considering an individual who prefers a location on the segment $\left(\overline{h}_v, \overline{h}_v + \Delta\overline{h}_v\right)$ without the notch and therefore prefers the cutoff $\overline{h}_v$ with the notch (conditional on buying). For such an individual, the change in the threshold fixed cost Δq^* induced by the notch is given by

$$\Delta q^* = u\left(\overline{c}, \overline{h}_v/p\right) - u\left(c^*, h^*\right), \tag{1.2.6}$$

where $\overline{c}, \overline{h}_v/p$ is the consumption bundle obtained at the notch. As the preferred point absent the notch h^* converges to the cutoff $\overline{h}_v/p$ from above (and hence c^* converges to $\overline{c}$), Δq^* converges to zero and extensive responses disappears. Intuitively, if in the absence of the notch, an individual would choose to buy a house slightly above $\overline{h}_v$, then in the presence of the notch, she will be better off by buying a house at $\overline{h}_v$ (which is almost as good) rather than not buying at all. This reasoning implies that extensive responses affect the density distribution as illustrated in Panel D of Figure 1. These effects can be summarized in the following proposition

PROPOSITION 1 (**Notches**). *A transaction tax featuring a notch at a property value $\bar{h}_v$ at which the proportional tax rate jumps from t to $t + \Delta t$ induces*

(i) *an* ***intensive margin response*** *as agents in a house price range $\left(\overline{h}_v, \overline{h}_v + \Delta\overline{h}_v\right)$ bunch at the threshold $\bar{h}_v$, where the width of the bunching segment $\Delta\overline{h}_v$ is monotonically increasing in the demand elasticity α as characterized by equation (1.2.9); and*

(ii) *an* ***extensive margin response*** *as agents in the house price range $\left(\overline{h}_v, \infty\right)$ who are sufficiently close to indifference between buying and not buying, $q \in (q^* + \Delta q^*, q^*)$, no longer buy. The extensive response converges to zero just above the cutoff as $\Delta q^* \to 0$ for $h_v \to \overline{h}_v^+$.*

These effects imply that around a cutoff at $\bar{h}_v$ the density of house values that we observe will feature an excess mass of

$$B\left(\bar{h}_v\right) = \int_{\bar{h}_v}^{\bar{h}_v + \Delta\bar{h}_v} g_0\left(h_v\right) dh_v \approx g_0\left(\bar{h}_v\right) \Delta\bar{h}_v, \tag{1.2.7}$$

where $B(\bar{h}_v)$ is excess mass at the cutoff and $g_0(h_v)$ is the counterfactual density of house values (i.e. the density that would prevail absent the notch). The approximation is accurate to the extent that the counterfactual is approximately uniform around the notch. Based on equation (1.2.7), it is possible to recover the house price response $\Delta\bar{h}_v$ based on estimates of the counterfactual distribution $g_0(h_v)$ and bunching $B(\bar{h}_v)$.

The relationship (1.2.7) implicitly assumes that there is just one bunching segment $(\bar{h}_v, \bar{h}_v + \Delta\bar{h}_v)$, which amounts to assuming that ε is homogeneous in the population. Our conceptual framework allows for heterogeneity and we can also account for it in the empirical implementation. There will be a price response $\Delta\bar{h}_v(\varepsilon)$ and a counterfactual density $\tilde{g}_0(h_v, \varepsilon)$ associated with each type ε. In this case, eq. (1.2.7) can be generalized to

$$B(\bar{h}_v) = \int_{\varepsilon}\int_{\bar{h}_v}^{\bar{h}_v+\Delta\bar{h}_v(\varepsilon)} \tilde{g}_0(h_v, \varepsilon)\, dh_v d\varepsilon \approx g_0(\bar{h}_v)\, E[\Delta\bar{h}_v], \tag{1.2.8}$$

where $E[\Delta\bar{h}_v]$ is the average price response across all ε. As before, the approximation requires that the counterfactual density is locally uniform in house prices h_v (but not type x) around the notch point. Equation (1.2.8) shows that estimates of the counterfactual distribution and bunching allows us to recover the average house price response in the population.

As shown by Kleven & Waseem (2013), the relationship between bunching and the demand elasticity can be characterized using two distinct approaches. The structural approach considers the marginal bunching individual who is indifferent between the notch point and her best interior location. This indifference condition along with the first-order condition for the no-notch location $\overline{h}_v + \Delta\overline{h}_v$ implies that the marginal bunching individual satisfies

$$\frac{1}{1+\Delta\bar{h}_v/\bar{h}_v} - \frac{1}{1+1/\varepsilon}\left[\frac{1}{1+\Delta\bar{h}_v/\bar{h}_v}\right]^{1+1/\varepsilon} - \frac{1}{1+\varepsilon}\left[1+\frac{\Delta t}{1+t}\right]^{1+\varepsilon} = 0. \tag{1.2.9}$$

With our estimate of the width of the bunching segment $\Delta\overline{h}_v$ and the tax parameters $\overline{h}_v$ and $\Delta t/(1+t)$, this condition gives a unique demand elasticity ε. However, since the structural approach relies heavily on the functional form for utility as well as the competitive market assumption, we follow Kleven & Waseem (2013) and also characterize the elasticity using a reduced-form approximation. This approach relates the house price response $\Delta\bar{h}_v$ to the change in the implicit marginal tax rate between $\bar{h}_v$ and $\bar{h}_v + \Delta\bar{h}_v$ created by the notch.

Defining this implicit marginal tax rate as $t^* = \left\{T\left(\bar{h}_v + \Delta\bar{h}_v\right) - T\left(\bar{h}_v\right)\right\}/\Delta\bar{h}_v$, the house price elasticity with respect to $(1+t^*)$ is given by

$$\varepsilon_r \equiv \frac{\Delta\bar{h}_v/\bar{h}_v}{\Delta t^*/\left(1+t^*\right)} \approx \frac{\left(\Delta\bar{h}_v/\bar{h}_v\right)^2}{\Delta t/\left(1+t\right)}, \tag{1.2.10}$$

where the notch-induced change in the implicit marginal tax rate is approximated as $\Delta t^* \approx \left\{\Delta t \cdot \bar{h}_v\right\}/\Delta\bar{h}_v$. The advantage of estimating a house price elasticity with respect to the *marginal* tax rate (using notches that create jumps in the *average* tax rate) is that it allows for an evaluation of house price responses in the interior of tax brackets (where individuals are responding to marginal tax rate changes) and also for an evaluation of alternative non-notched tax structures.

1.2.2. A Dynamic Extension to the Model. To guide the empirical analysis of temporary stimulus policy, let us briefly consider a dynamic extension of the previous model. In general, temporary tax changes will create both timing responses and extensive margin responses in the housing market. To see this, consider a simple two-period extension of the model in which agents maximize lifetime utility $u_1\left(c_1, h_1\right) + \beta u_2\left(c_2, h_2\right)$ where the per-period utility functions are given by

$$u_s\left(c_s, h_s\right) = c_s + \frac{A_s}{1+1/\varepsilon_s}\left(\frac{h_s}{A_s}\right)^{1+1/\varepsilon_s} - q_s \cdot I\left\{h_s \neq h_{s-1}\right\} \tag{1.2.11}$$

Note that all the preference parameters $\{A_s, \varepsilon_s, q_s\}$ are allowed to vary between periods. In each period, agents choose whether to be active in the housing market or whether to remain in their current house (either a rented house or a house they purchased in a previous period). For simplicity we will assume that all agents start out renting so that $h_0 = 0$ for all agents, but this does not affect any of the results. If agents choose to be active in the housing market in period s they pay a fixed cost q_s, choose the amount of housing to purchase h_s, and if $h_{s-1} \neq 0$, they also simultaneously sell their existing house. Agents also receive income of y_s in each period and so face a budget constraint analogous to equation (1.2.2) in each period $s \in \{1, 2\}$ given by[6]

$$c_s + p_s \left[(1 + t_s) h_s - h_{s-1} \right] \cdot I \{ h_s \neq h_{s-1} \} = y_s \tag{1.2.12}$$

Solving the model backwards, consider an individual who enters period 2 with housing $h_1 \geq 0$. Just as in the static case, this individual will maximize $u_2 (c_2, h_2)$ subject to her budget constraint (1.2.12) and, conditional on buying, demand housing $h_2^* = A_2 \left[(1 + t_2) p_2 \right]^{\varepsilon_2}$. This agent therefore buys a new house iff $u_2 (c_2^*, h_2^*) > u_2 (y_2, h_1)$ and we can write her indirect utility as $v_2 ((1 + t_2) p_2, y_2, h_1) = \max \{ u_2 (c_2^*, h_2^*), u_2 (y_2, h_1) \}$. Working backwards, individuals in period 1 anticipate the effect that their housing choices will have on their utility in period 2, so they maximize $u_1 (c_1, h_1) + v_2 ((1 + t_2) p_2, y_2, h_1)$ subject to the period 1 budget constraint (1.2.12), again yielding a period-1 housing demand function h_1^* conditional upon buying. Individuals therefore buy in period 1 whenever $u_1 (c_1^*, h_1^*) + \beta v_2 ((1 + t_2) p_2, y_2, h_1^*) > u_1 (y_1, 0) + \beta v_2 ((1 + t_2) p_2, y_2, 0)$. In this model there will, in general, be four groups of agents: those who buy a house in period 1 and stay in it in period 2; those who buy in period 1 and then move in period 2; those who do not buy in period 1 but do so in period 2; and those who never buy.

If we now consider a reduction in the first-period tax t_1, this unambiguously makes buying a house in period 1 more attractive by lowering the net-of-tax price of housing. This has two conceptual effects on the level of activity in the housing market in period 1. First there will be a *timing effect* as agents who were close to indifferent between buying in period 2 and buying in period 1, i.e. those for whom $y_1 + \beta u_2 (y_2 - p_2 (1 + t_2) h_2^*, h_2^*) \approx u_1 (c_1^*, h_1^*) + \beta u_2 (y_2, h_1^*)$, buy a house in period 1 instead of waiting until period 2. Second, there will be an *extensive margin effect* by two types of agents. Those who were close to indifferent between never buying and buying in period 1, i.e. those for whom $y_1 + \beta u_2 (y_2, 0) \approx u_1 (c_1^*, h_1^*) + \beta v_2 ((1 + t_2) p_2, y_2, h_1^*)$, buy in period 1 instead of not buying at all. Furthermore, those who were close to indifferent between buying only in period 2 and buying in both periods, i.e. those for whom $y_1 + \beta u_2 (y_2 - p_2 (1 + t_2) h_2^*, h_2^*) \approx$

[6] In this formulation, we can think of p_s as the price of 1 unit of housing services in every period from the current period onwards. In a model without liquidity constraints and in which utility is quasilinear this is, of course, immaterial. Moreover, even in a richer model the qualitative predictions that we explore in our empirical analysis will be unchanged.

$u_1(c_1^*, h_1^*) + \beta u_2(y_2 - p_2[(1+t_2)h_2^* - h_1^*], h_2^*)$, are induced to buy twice over their lifetime instead of only once. To summarize,

PROPOSITION 2 (**Temporary Stimulus**). *An unanticipated temporary stimulus policy reducing the transaction tax in period 1, but not in period 2, causes*

(i) *a* ***timing effect*** *as agents who were sufficiently close to indifference between buying in period 1 and buying in period 2 (preferring the latter) are induced to shift their house purchase forward; and*

(ii) *an* ***extensive margin effect*** *by two sets of agents. Those who were sufficiently close to indifference between buying in period 1 and never buying (preferring the latter) are induced to buy in period 1. Those who were sufficiently close to indifference between buying in both periods and buying only in period 2 (preferring the latter) are induced to buy twice over their lifetime intead of only once.*

1.2.3. Revenue and Welfare Effects of a Property Transaction Tax. In this section we consider the revenue and welfare consequences of property transaction taxes, summarising the effects in two key statistics. First, we derive the marginal cost of public funds, defined as the marginal welfare cost per dollar of revenue collected through the transaction tax (see e.g., Kleven & Kreiner 2006; Dahlby 2008). Second, we derive the revenue-maximizing tax rate (the Laffer rate). Tax rates beyond the Laffer rate reduce revenue as well as the utility of each homebuyer, and are therefore Pareto inefficient.

Since here we are interested in the long run welfare properties of the tax, we return to the static version of the model. We focus on a proportional tax rate, which simplifies the analysis and is sufficient for (approximately) evaluating the UK transaction tax. This is due to the fact that the UK system is proportional *within* brackets and that notches at bracket cutoffs create only *local* distortions that will in general not contribute substantially to total revenue and welfare (though they are very useful for identification). Hence we will be able to apply our measure of the marginal cost of public funds separately to each bracket in order to evaluate the welfare effects through the distribution.

Aggregate welfare in our simple economy is given by

$$W = \int_A \int_\varepsilon \int_0^{q^*} u(c^*, h^*) f(A, \varepsilon, q)\, dq d\varepsilon dA + \int_A \int_\varepsilon \int_{q^*}^{\infty} u(y, 0) f(A, \varepsilon, q)\, dq d\varepsilon dA$$

where we are abstracting from distributional considerations by weighting all individuals equally (though these could easily be incorporated), and tax revenues are

$$T = th_v^m$$

where $h_v^m = \int_A \int_\varepsilon \int_0^{q^*} ph^* f(A, \varepsilon, q)\, dq d\varepsilon dA$ is the average house value. If we consider a small increase in the tax rate by dt, then we can define the marginal cost of funds of the additional revenues raised as

$$MCF = -\frac{dW}{dT} \tag{1.2.13}$$

where dW is the change in welfare induced by the change dt, and dT are the additional revenues raised.

Turning first to the effect on tax revenues, the tax rate increase has three effects. First, the tax increase has a *mechanical effect* from taxing all transactions at a higher rate, raising additional revenues of $dM = h_v^m dt$. Second, there is an *intensive effect* as individuals buy less housing, conditional on buying. This effect reduces revenues by an amount $dI = \frac{t}{1+t}\bar{\varepsilon} h_v^m dt$, where we define $\bar{\varepsilon} = -\int_A \int_\varepsilon \int_0^{q^*} \varepsilon \frac{ph^*}{h_v^m} f(A, \varepsilon, q)\, dq d\varepsilon dA$ as the house-value weighted average demand elasticity in absolute value. Finally, the tax increase has an *extensive effect* as some individuals no longer find it worthwhile to buy a house. Defining the extensive margin elasticity in absolute value as $\eta = -\frac{\partial F(q^*|A,\varepsilon)dq}{\partial(1+t)} \frac{1+t}{F(q^*|A,\varepsilon)}$ the extensive effect reduces revenues by $dE = \frac{t}{1+t}\bar{\eta} h_v^m dt$ where $\bar{\eta} = \int_A \int_\varepsilon \eta \frac{ph^*}{h_v^m} f(A, \varepsilon)\, d\varepsilon dA$ is the house-value weighted average extensive margin elasticity.

Since all individuals are optimizing their choice of housing demand h^*, an envelope theorem argument implies that we only need to consider the direct effects of the tax increase on welfare, so that

$$dW = -\int_A \int_\varepsilon \int_0^{q^*} \frac{\partial c^*}{\partial t} dt = -h_v^m dt = -dM$$

Inserting these components into (1.2.13), we obtain

$$MCF = \frac{dM}{dM - dI - dE} = \frac{1}{1 - \frac{t}{1+t}(\bar{\varepsilon} + \bar{\eta})} \tag{1.2.14}$$

The second summary statistic we consider is the Laffer rate, defined as the rate t_L at which $dT = 0$. Using the above expressions, we obtain that

$$\begin{aligned} dT &= dM - dI - dE = 0 \\ t_L &= \frac{1}{\bar{\varepsilon} + \bar{\eta} - 1} \end{aligned} \tag{1.2.15}$$

and we note that the marginal cost of funds at the Laffer rate is infinite. As mentioned above, the Laffer rate is a key statistic as it represents the Pareto bound of a tax system.

1.3. Context and Data

1.3.1. The UK Property Transaction Tax: Notches and Reforms. The UK property transaction tax—Stamp Duty Land Tax (SDLT)—is imposed on the transaction value of land and any construction on the land, known as the "chargeable consideration".[7] This is defined in the broadest possible terms to include anything of economic value given in exchange for land or property, including money, goods, works or services, and transfers of debts. The statutory incidence of the SDLT falls on the buyer, who is required to file a stamp duty return and remit tax liability to HMRC within a few weeks of the completed transaction. The SDLT is a significant source of government revenue in the UK, much more so than other wealth transfer taxes such as inheritance taxation and capital gains taxation. The SDLT has raised revenue of around 0.6% of GDP over recent years,[8] and the political debate in the UK suggests that future rates (on highly priced properties) are more likely to go up than down.

A central aspect of the stamp duty is that it features discrete jumps in tax liability—notches—at threshold property prices. Tax liability is calculated as a proportional tax rate times the transacted property price, with different tax rates in different price brackets. Hence, as the purchase price crosses a bracket threshold, a higher tax rate applies to the entire amount and not just the portion that falls above the cutoff as in standard graduated schedules. Figure 2 illustrates the stamp duty schedule for residential property in tax year

[7]The chargeable consideration includes the buildings and structures on the land as well as fixtures and fittings (such as in bathrooms and kitchens), but excludes freestanding furniture, carpets or curtains. If such extras are included in the sale, the buyer and seller are to agree on the market value of these extras and subtract it from the chargeable consideration. See http://www.hmrc.gov.uk/sdlt/calculate/value.htm for details.

[8]See http://www.hmrc.gov.uk/stats/tax_receipts/tax-receipts-and-taxpayers.pdf.

2012–13.[9] The schedule features five notches as the proportional tax rate jumps from zero to 1% at a price of £125,000, from 1% to 3% at a price of £250,000, from 3% to 4% at a price of £500,000, from 4% to 5% at a price of £1,000,000, and finally from 5% to 7% at a price of £2,000,000.[10] The schedule is different for residential property in certain disadvantaged areas (where the first bracket threshold is at a higher price) as well as for non-residential property. It is worth noting that the buyer cannot mortgage the SDLT liability, it must be financed from savings, and so we should expect the SDLT to have large effects on liquidity constrained buyers. It should also be noted that stamp duty schedules are not indexed for inflation, which creates "bracket creep" as property price inflation pushes houses into higher stamp duty brackets.

Another important aspect of the stamp duty is that it has been subject to a great deal of policy experimentation over the years. As shown in Table 1, the main policy experiments during our data period have been (*i*) changes in the location of the lower notch and (*ii*) the introduction of new notches at £1,000,000 in April 2011 and at £2,000,000 in March 2012. It is worth describing the specific features of some of those policy changes as they will be important for the empirical analysis.

For the lower notch, the most salient change was the so-called *stamp duty holiday* between 3 September 2008 and 31 December 2009, which moved the first notch point from £125,000 to £175,000 and thereby eliminated stamp duty in a £50,000 range. The motivation of the program was to provide housing stimulus during the current recession.[11] The following features of the stamp duty holiday are important for our analysis. First, the beginning of the holiday was *unanticipated* as it was announced suddenly by the then Chancellor Alistair Darling on the day before its introduction. Although there was some media speculation about the possibility of a stamp duty holiday in the month leading up to the announcement, the details and start date of such a holiday were unknown. Second, the end of this holiday was *anticipated*. The initial announcement was that the holiday would last for one year (until September 2009), but in April 2009 this was extended until

[9] The UK tax year for personal taxes runs from April 6 in one year to April 5 the next year.

[10] At the £2,000,000 notch, the stamp duty rate jumps to 15% if the residential dwelling is purchased by certain "non-natural persons" such as corporations and collective investment schemes.

[11] Another stimulus program was implemented specifically for first-time buyers between 25 March 2010 and 24 March 2012. This program temporarily abolished the notch at £125,000, thereby eliminating stamp duty in the range between £125,000 and £250,000 for first-time buyers.

the end of 2009 and the government committed to no further extensions (and indeed did not grant any extensions). The sudden announcement of the stamp duty holiday and the preannounced committment to its end date allow us to compare the effects of expected and unexpected changes in tax policy. In particular, the pre-announced end date creates a *time notch* (a discrete jump in tax liability at a cutoff date) allowing us to analyze short-term timing effects. Finally, as the stamp duty holiday applied only to properties in a certain price range, we are able to study the stimulus effects of the policy and subsequent reversal (medium-term timing) using a difference-in-differences approach.

For the top notches, the introduction of a higher stamp duty rate above £1,000,000 was pre-announced a full year in advance, while the higher stamp duty rate above £2,000,000 was confirmed just one day before it took effect. Hence, the introduction of the £1,000,000 price notch (but not the £2,000,000 price notch) also creates a time notch that allows us to study anticipatory behavior.

The UK stamp duty appears to be characterized by relatively high compliance. According to HMRC estimates, the so-called *tax gap*—the difference between taxes owed and taxes paid on a timely basis—is between 4–5% of true stamp duty tax liability. This is lower than the tax gap estimates for most other taxes in the UK. It is perhaps not surprising that tax evasion is a minor issue for this tax when considering the following points. First, almost all property transactions in the UK are facilitated by licensed real estate agencies, implying that stamp duty tax evasion requires collusion between a buyer, a seller and a real estate agency (typically with multiple employees). Such evasion collusion involving many agents is unlikely to be sustainable (Kleven *et al.* 2009). Second, the scope for tax evasion is further reduced by the existence of a considerable lag between agreeing on a house price and completing the contract.[12] If the house price reported to tax authorities is lower than the true house price, the buyer must make a side payment to the seller. If the buyer makes the side payment at the time of agreeing on the house price, the seller would be able to renege before completing the contract and it would be difficult for the buyer to recoup the payment. If instead the buyer promises to make the payment at the time of completing the contract, the seller would take his property off the market with no credible commitment from the buyer that he would not renege later when the bargaining position of the seller

[12] This lag is about 2 months on average in the UK housing market (Besley *et al.*, 2011).

may be weaker. Hence, such side payments would be associated with substantial risk for either the buyer or the seller or both. Finally, as described above, the tax base is defined in an very comprehensive manner meaning that the scope for shifting or re-classification of specific features of the property to avoid the tax is limited. The one exception is the exclusion in the tax base of freestanding "extras" such as furniture and curtains. If such extras are included in the sale, the buyer and seller are to agree on the market value of these extras and subtract it from the chargeable consideration, which creates an opportunity to evade stamp duty by overvaluing such items (while undervaluing the rest of the property by the same amount). However, reporting large amounts of tax exempt extras is an audit trigger, limiting the degree to which such behavior is possible. For all of these reasons, we believe that house prices reported on stamp duty tax returns reflect true house prices in most (but not all) cases.

1.3.2. Data and Raw Time Series Evidence. The empirical analysis is based on administrative data covering the universe of stamp duty (SDLT) returns in the UK from November 2004 to October 2012. Since most property transactions require the filing of an SDLT return (the main filing exemption being for property transactions under £40,000), our data is close to the universe of property transactions in the UK. The full dataset contains about 10 million transactions. The dataset contains rich tax return information for each transaction, but currently very little information outside the tax return.

The housing market has seen substantial turmoil during the period we consider. Figure 3 shows the monthly number of house transactions (Panel A) and the monthly average property price (Panel B) in all of the UK and in London alone. The figure shows nominal prices (real prices give the same qualitative picture) and normalizes both the price and the number of transactions to one at the start of the period. We make the following observations. First, housing market activity collapses between late 2007 and early 2009 as the the number of transactions falls by around two-thirds. There has been some recent recovery, but activity is still very far from pre-recession levels. Second, property prices also fall between late 2007 and early 2009, but the price drop is less dramatic and the subsequent recovery much stronger. Third, property prices (though not activity) in London have evolved differently than in the rest of the UK during the recession. While UK-wide property prices have recovered only partially in the past couple of years, London property prices are almost back on their pre-recession trend. Fourth, the recovery in house prices and activity throughout 2009 coincides with the stamp duty holiday, which has been used as an argument that the policy had the desired effect. We will take a quasi-experimental approach to evaluate how much of the recovery (if any) can indeed be explained by the stamp duty holiday. Finally, average house prices in London feature a sharp spike in early 2011 and a subsequent dip, which constitutes our first piece of evidence of a behavioral response to stamp duty incentives. This spike reflects excess trading of houses above £1,000,000 just before the pre-announced introduction of the £1,000K stamp duty notch on 6 April 2011 and the dip reflects missing trading of such houses just after the introduction of the notch—a short-term timing response to an anticipated tax change.

1.4. House Price Responses to Transaction Taxes: Notches

1.4.1. House Price Responses to Static Notches. This section presents static results using price notches during periods when they are stable. We consider residential property transactions that incur a stamp duty land tax liability.[13] Figure 4 considers the two notches located at cutoff prices of £250,000 (Panel A) and £500,000 (Panel B), both of which have remained in place throughout the period of our data. Each panel shows the empirical distribution of house values (blue dots) as a histogram in £5,000 bins and an estimated counterfactual distribution (red line). Following Chetty *et al.* (2011) and Kleven & Waseem (2013), the counterfactual distribution is estimated by fitting a flexible polynomial to the empirical distribution, excluding data in a range around the notch, and allowing for round-number fixed effects to capture rounding in the price data.[14] The excluded range

[13] Results for non-residential property are qualitatively similar, but noisier as we have far fewer observations.

[14] Grouping transactions into price bins of £100, the regression used to estimate the counterfactual distribution around a notch at price $\bar{h}_v$ is given by

$$c_i = \sum_{j=0}^{q} \beta_j (z_i)^j + \sum_{r\in\mathcal{R}} \eta_r I\left\{\frac{\bar{h}_v + z_i}{r} \in \mathbb{N}\right\} + \sum_{k=\bar{h}_v^-}^{\bar{h}_v^+} \gamma_k I\{i = k\} + \mu_i, \tag{1.4.1}$$

where c_i is the number of transactions in price bin i, z_i is the distance between price bin i and the cutoff $\bar{h}_v$, and q is the order of the polynomial ($q = 5$ in Figure 4). The second term in (1.4.1) includes fixed effects for prices that are multiples of the round numbers in the set $\mathcal{R}$, where $\mathcal{R} = \{500, 1000, 5000, 10000, 25000\}$, $\mathbb{N}$ is the set of natural numbers, and $I\{\cdot\}$ is an indicator function. Finally, the third term in (1.4.1) excludes a region $\left(h_v^-, h_v^+\right)$ around the notch that is distorted by bunching responses to the notch, and μ_i is a residual reflecting misspecification of the density equation. Our estimate of the counterfactual distribution is defined as the predicted bin counts $\hat{c}_i$ from (1.4.1) omitting the contribution of the dummies in the excluded range, and excess bunching is estimated as the difference between the observed and counterfactual bin counts in the part of the excluded range that falls below the notch $\hat{B} = \sum_{i=\bar{h}_v^-}^{\bar{h}_v} (c_i - \hat{c}_i)$. We may also define an estimate of is demarcated by vertical dashed lines; the lower bound is set at the point where excess bunching starts and the upper bound is set at the point where the hole ends (where the empirical distribution above the cutoff changes slope from positive to negative).

As discussed in detail by Kleven & Waseem (2013), due to the presence of potential extensive responses above the excluded range, this estimation procedure intends to provide a "partial counterfactual" stripped of intensive responses, but not extensive responses. This partial counterfactual corresponds to the border of the light-gray area in Panel D of Figure 1, which is smooth around the cutoff. To simplify, our estimation of the counterfactual distribution ignores the marginal shift in the distribution above the hole due to intensive responses in the interior of the upper bracket. It is feasible to account for this shift in the distribution when estimating the counterfactual,[15] but given the size of the incentive (a marginal tax rate change of 1–2% above the notch) and the house price elasticities that we find, this shift will be extremely small and have no substantive effect on any of our conclusions.

In Figure 4, each panel shows estimates of excess bunching below the notch scaled by the counterfactual frequency at the notch (b), the size of the hole (missing mass) above the notch scaled by the counterfactual frequency at the notch (m), the difference between these two ($m - b$), the average house price response to the notch (Δh_v), and the tax liability change at the notch (ΔTax). Our main findings are the following. First, both notches create large and sharp bunching below the cutoff. Excess bunching is 1.85 and 1.64 times the height of the counterfactual distribution at £250,000 and £500,000, respectively, and is strongly significant in each case. Second, both notches are associated with a large hole in the distribution above the cutoff. The size of the hole is larger than the size of excess bunching, although the difference between the two is not statistically significant from zero. Third, the hole in the distribution spans a £25,000 range above each cutoff, implying that the most responsive agents reduce their transacted house value by five times as much as

missing mass (the hole) above the notch as $\hat{M} = \sum_{i > \bar{h}_v}^{\bar{h}_v^+} (\hat{c}_i - c_i)$, but this statistic is not used in the estimation of house price responses and house price elasticities (see section 1.2.1). Standard errors on all estimates are calculated based on a bootstrap procedure as in Chetty *et al.* (2011). As a robustness check we have tried values between 4 and 7 for the order of the polynomial and our results are not significantly altered.

[15]This can be done by using an initial estimate of the house price elasticity (based on ignoring the shift in the upper distribution) to obtain an initial estimate of the distribution shift, re-estimate the counterfactual and the house price elasticity to respect the initial estimate of the distribution shift, and continue the procedure until the estimation converges.

the jump in tax liability of £5,000.[16] Fourth, the average house price response is £10,000 at both the £250,000 notch and the £500,000 notch, a response that is twice as large as the tax jump.

We now turn to the lower notch, the location of which has changed several times during the period under consideration. The cutoff was located at £60,000 until 16 March 2005, at £120,000 between 17 March 2005 and 22 March 2006, at £125,000 between 23 March 2006 and 2 September 2008, at £175,000 between 3 September 2008 and 31 December 2009, and again at £125,000 from 1 January 2010 onwards. This section takes a static approach by considering bunching responses within each of these five periods separately, while the next section investigates dynamic adjustment paths around the reform episodes. Figure 5 shows results for the five periods in separate panels, each of which is constructed as in the Figure 4. The findings for the lower notch are qualitatively consistent with those for the other notches, with a clear and statistically significant bunching response to the tax notch in each period. The size of the bunch and the hole is smaller at the lower notch than at the upper notches, but so is the size of the notch. The effect of the notch on the average transacted house value is between £3,500 and £5,000, or about 4–5 times the size of the tax liability jump so responses are actually proportionally larger at the bottom.

To facilitate comparison across notches, and to assess house price responses outside the regions around notches (where individuals are responding to standard marginal tax incentives), Table 2 converts the house price responses at each of the notches into house price elasticities with respect to the marginal tax rate using both the structural and reduced-form approaches outlined in section 1.2.1. Despite the large house price responses, the elasticities are relatively modest due to the enormous marginal tax rate variation driving those responses. The elasticity declines monotonically through the price distribution, ranging from around 0.2–0.3 at the bottom of the distribution to below 0.05 at the top. As shown by Kleven & Waseem (2013), the reduced-form approach in general over-estimates the elasticity, and consistent with this, we find that the reduced-form elasticity is slightly

[16]This finding is interesting when considering mortgage terms in the UK. Mortgage rates depend on the downpayment as a share of the house price according to a notched schedule, with the credit terms improving drastically if the borrower is able to put down a deposit of at least 20%. Hence, if a buyer is targeting the 20% mortgage notch *and* is liquidity constrained, the house price is fixed at five times savings net of stamp duty payments (recall that stamp duty cannot be mortgaged). This implies that the house price responds precisely by a factor of five to the stamp duty. In future work, we plan to investigate the role of liquidity constraints for the joint responsiveness to taxes and mortgage rates using administrative mortgage data.

larger than the structural elasticity at each notch. These modest house price elasticities with respect to the marginal tax rate imply that house price responses outside the regions around notches are quite modest, though larger at the bottom.

In 2011 and 2012, the government introduced two new notches affecting very high value properties, one at £1 million on 6 April 2011 and another one at £2 million on 22 March 2012. The stamp duty notch at £2 million is commonly referred to as the "mansion tax". Even though these are very recent notches, they have already created a clear house price distortion as shown in Appendix Figure 15. This figure is constructed in the same way as the previous ones, except that the counterfactual distribution is obtained differently. We take advantage of the tax reform (notch introduction) by comparing the empirical house price distribution after the introduction of the notch to the empirical distribution in the year leading up to the introduction of the notch. The results are qualitatively very similar to the previous results, with an average house price response of £30,000 at the £1 million notch (3 times the tax liability jump of £10,000) and £100,000 at the £2 million notch (2.5 times the tax liability jump of £40,000).

Finally, when interpreting our results, note that reported house prices in our data can be described by $h_v \equiv p \cdot h - e$, where p is the price per unit of quality-adjusted housing, h is the amount of quality-adjusted housing, and e is stamp duty evasion. This means that, in general, our estimates of house price responses combine price changes Δp (incidence), real demand changes Δh (buying a lower-quality house), and evasion responses Δe. As clarifed in the theory section, the price incidence effect reflects potential match-specific price bargaining rather than standard market-level incidence driven by aggregate demand and supply (which does not by itself create bunching). Our estimates of house price responses are conceptually similar to the estimation of taxable income responses (e.g. Saez *et al.* 2012), which combines wage bargaining effects, real labor supply, and evasion.

1.4.2. House Price Responses to Moving Notches. This section investigates the dynamics of behavioral adjustment to the changes in the position of the lower notch that were mentioned above. When considering dynamic adjustments, it is important to keep in mind that there is always a lag between agreeing on a purchase price and completing the housing contract. In the UK housing market, this lag is under 90 days for most transactions and about 60 days on average (Besley *et al.* 2011). Since the official transaction date in our data refers to contract completion, the time it takes for the market to settle into a new equilibrium is bounded from below by about 3 months.

Figure 6 considers the movement of the lower notch from £120,000 to £125,000 on 23 March 2006. Each panel shows the empirical and counterfactual distributions in a given month between February 2006 and September 2006. The two vertical lines demarcate the £120,000 and £125,000 cutoffs and are either solid green (for the cutoff that is active in month in question) or dashed black (for the cutoff that is inactive). April 2006 is the first full month where the new cutoff is in place. The figure shows very clearly how the bunch moves over time in response to the changed location of the notch. Most of the adjustment has occurred after four months (in July 2006) and a new equilibrium has been reached after 6 months (in September 2006).[17] Hence, most of the lag in the adjustment to the new equilibrium can be explained by the administrative lag between contract exchange and contract completion.

The next three figures consider the movement of the lower notch from £125,000 to £175,000 on 3 September 2008 (the start of stamp duty holiday) and the subsequent movement back to £125,000 on 1 January 2010 (at the end of stamp duty holiday). When interpreting the findings, it is worth keeping in mind that the start of the holiday was unanticipated while the end of the holiday was anticipated (see section 1.3.1). Figure 7 shows monthly bunching graphs over a 12-month period around the beginning of the holiday. It is constructed like the preceding figure, except that we now add estimates of excess bunching b around the two cutoffs in each month.[18] The main findings are the following. First, it takes 3–4 months for bunching at the old £125,000 cutoff to disappear (bunching becomes statistically insignificant for the first time in December 2008), corresponding roughly to the lag between contract agreement and completion. Second, it takes about 3 months for bunching at the new £175,000 cutoff to build up and reach a steady state (bunching b is around 0.9 from November 2008 onwards). Third, although bunching at £175,000 in the winter months of 2008/09 is smaller in absolute terms than bunching at £125,000 in the summer months before the holiday, bunching in proportion to

17 Animated versions of all the figures from this section that show the dynamics more vividly can be found at http://personal.lse.ac.uk/kleven/Downloads/MyPapers/WorkingPapers/best-kleven_landnotches_april2013_videos.pdf

18 Animated versions of these figures are online at the address in footnote 17.

the counterfactual distribution (b)—the right measure of responsiveness—is in fact slightly larger at £175,000. The presence of smaller absolute bunching at £175,000 is a result of seasonality in the housing market with fewer house transactions in the winter than in the summer.[19] The presence of larger relative bunching b at £175,000 is consistent with the fact that this notch is larger than the previous one at £125,000 (tax liability jumps of £1,750 and £1,250 respectively).

Figure 8 turns to the 12-month period around the end of the holiday on 1 January 2010 and is constructed exactly as the preceding figure. It is interesting to see the difference in the speed of adjustment to a tax change that is fully anticipated. First, the bunching at £175,000 vanishes immediately in January of 2010 when this cutoff is no longer a notch point. This shows that buyers and sellers did indeed anticipate the end of the holiday and made sure to complete their housing contracts before the end of December 2009. We see such behavior in the graph for December 2009: there is a large upward shift in the December distribution between £125,000 and £175,000 (even though this is normally a low-season month) and an increase in excess bunching at £175,000. The next section investigates such short-term timing behavior in greater detail. Second, it takes about 2 months for bunching at the new £125,000 cutoff to build up and reach a stable equilibrium (b is roughly constant from February 2010 onwards). While this is faster adjustment than at the start of the holiday, it is not as fast as the disappearance of bunching at the end of the holiday. The implication is that, while buyers and sellers were rushing to *complete* agreed housing contracts below the the £175,000 notch just before the end of the holiday (immediate disappearance of old bunching), they did not to the same degree *agree* (but not complete) housing contracts below the £125,000 notch just before the end of the holiday (slower emergence of new bunching).

Figure 9 summarizes the evidence in the preceding figures by showing the monthly bunching estimate b from January 2007 to January 2011 at the £125,000 cutoff (blue dots) and the £175,000 cutoff (orange crosses) with 95% confidence intervals around each series. The solid vertical lines demarcate the beginning and end of the stamp duty holiday, while the dashed vertical line demarcates the de facto time at which the holiday took full effect given the lag between agreed and completed house purchases. The figure highlights just

[19] Seasonality in the housing market is a well-known phenomenon that has been studied in the macro literature (e.g. Ngai & Tenreyro 2012).

how sharply house prices react to tax notches and to changes in tax notches even at the monthly level. The level of bunching at the £125,000 cutoff is remarkably constant on each side of the holiday, while the level of bunching at the £175,000 cutoff is constant during the holiday. The steady state level of bunching at £175,000 ($b \approx 0.9$) is larger than at £125,000 ($b \approx 0.6$) as the former notch is larger. Once we account for the built-in sluggishness due to the time it takes to complete a housing contract, the market adjusts to a new stable equilibrium remarkably quickly. We also do not see any difference in price responsiveness during good times and bad times (compare early part of 2007 to the rest of the period).

Compared to recent bunching evidence from labor markets (e.g. Saez 2010; Chetty *et al.* 2011; Kleven & Waseem 2013), the remarkable sharpness of our evidence suggests that behavioral responses in the housing market are much less affected by optimization frictions such as inattention, inertia, etc. Our evidence suggests that agents in the housing market respond precisely and quickly to tax incentives.

1.5. Timing and Extensive Margin Responses: Tax Reforms and Stimulus

We saw in the previous section that house prices respond sharply to the transaction tax, and adjust quickly when the notches in the schedule move around. In this section, we investigate the effect of the transaction tax on whether individuals transact, and if so, when they choose to transact. Then, with these estimates, we evaluate the revenue and welfare consequences of the transaction tax in the long run, and the efficacy of using transaction tax cuts as a stimulus policy.

1.5.1. Short Term Timing Responses to Anticipated Tax Changes. As described in section 1.3.1, the tax increase at the end of the stamp duty holiday was fully anticipated, creating a *time notch* on 1 January 2010 for houses between £125,000 and £175,000. This time notch creates a strong incentive for individuals to conclude their transactions before New Year, and bunching of the timing of transactions allows us to estimate this short-term timing response.

Before discussing the empirical results, we make two remarks. First, the housing market almost shuts down between Christmas and New Year, so the notch is effectively a notch just before Christmas. Hence, agents should respond to the notch by moving the date of purchase from the early weeks of 2010 to the third week of December 2009. Second, the existence of the Christmas holiday (with or without a tax notch) may in itself lead to a piling up of house transactions in the third week of December. This means that we cannot analyze the time notch using a "pure" bunching strategy as observed bunching in transactions before Christmas 2009 may overstate the response to the tax notch. We therefore pursue a difference-in-bunching strategy by comparing bunching in the treated group (transactions between £125,000–£175,000 in December 2009) to bunching in control groups (other years and/or other price ranges).

Figure 10 shows the weekly number of transactions around New Year in different price ranges and different years. Panel A compares the treated price range £125,000–£175,000 in the treated period 2009/10 to surrounding price ranges in the same period. The treated group features very strong bunching just before the notch and a large hole after the notch. The control groups also feature bunching and a hole (Christmas effect), but to a much smaller extent. Furthermore, the shutdown of activity between Christmas and New Year is less extreme in the treated group than in the control groups.

To evaluate the timing response, we estimate excess bunching in each distribution during the last three weeks of the year using a bunching approach analogous to our approach for the price notches.[20] The timing response is then given by the difference between bunching in the treated range and average bunching in the surrounding control ranges (D-i-Bunching in the figure). We find that excess mass induced by the time notch is almost 3 times the height of the counterfactual and strongly significant, implying that the average timing response to the notch is 3 weeks. Panel B is constructed in the same way, except that it compares the treated price range £125,000–£175,000 in the treated period 2009/10 to the same price range in other periods (one year earlier or two years earlier). The results are very similar, with estimated excess mass before the notch being somewhat larger and still strongly significant. The placebo tests in the bottom panels repeat the strategy in Panel A (comparing different price ranges), but one year or two years earlier. In each case,

[20] In particular, we estimate

$$c_w = \sum_{j=0}^{7} \beta_j (z_w)^j + \eta I\{w \in \text{end of month}\} + \sum_{k=\bar{w}^-}^{\bar{w}^+} \gamma_k I\{w = k\} + \mu_w, \tag{1.5.1}$$

where c_w is the number of transactions in week w and z_w is the distance of week w from the end of 2009. The second term is a fixed effect for weeks at the end of the month (which feature heavier trading in every month), while the third term excludes weeks in a range $(\bar{w}^-, \bar{w}^+)$ which we set to include the last 3 weeks of 2009 and the first 10 weeks of 2010.

the timing effect is close to zero and statistically insignificant. Overall, this provides very compelling evidence of short-term timing responses to anticipated tax changes, consistent with the sharpness of price responses discussed above. These findings contribute to the previous literature on the timing of the realisation of taxable income (Auerbach, 1988; Burman & Randolph, 1994; Goolsbee, 2000).

1.5.2. Medium Term Timing and Extensive Margin Responses to Stimulus. The stamp duty holiday was an unanticipated stimulus program with a fixed and fully anticipated end date. In the context of the dynamic model in section 1.2.2, this corresponds to an unanticipated tax cut in period s with no tax changes after period s, and in Proposition 2 we demonstrated that such a policy change has two conceptual effects on the level of activity in the housing market. First, there will be a *timing effect* as some agents who would have transacted a house after period s bring that transaction forward to period s. Second, there will be an *extensive margin effect* as some agents engage in additional house transactions over their lifetime, including house purchases in period s by those who would otherwise never buy (renter/homeowner margin) and house purchases in period s by those who continue to transact as often as they otherwise would have in other periods (more moving by existing homeowners). Hence, to assess the long-run impact of the transaction tax and to evaluate fiscal stimulus programs of this kind, it is crucial to obtain estimates not just of the total stimulus effect during the program (timing and extensive margin effects), but also of the degree to which it is driven by timing (all of which will be reversed after program withdrawal) and the length of the horizon over which there is re-timing (which determines the speed of reversal). This section provides compelling evidence on all three questions.

The stamp duty holiday temporarily cut the tax rate from 1% to 0% in the price range £125,000 to £175,000 without changing the tax rate in neighbouring price ranges, presenting us with an ideal opportunity to pursue a difference-in-differences approach. A naïve first cut at this (that we refine shortly) is to compare the evolution over time in transaction volumes in the treated range £125,000–£175,000 to a nearby control range. This is done in Figure 11, which compares the log monthly number of transactions in the treated range £125,000–£175,000 (blue dots) to a control range defined as £175,000–£225,000 (orange crosses). We have normalized the log number of transactions in each month by subtracting the average log number of transactions in the pre-treatment period (the 2 years leading up to the holiday) in order to make visual comparison of the two series easier. The solid vertical lines mark the beginning (3 September 2008) and the end (31 December 2009) of the stamp duty holiday.[21]

The two series display completely parallel trends leading up to the holiday and then begin to diverge precisely when the holiday starts. The positive effect of housing stimulus in the treated range increases during the first months of the holiday and features a sharp spike in the last month as people rushed to take advantage of the stimulus before it expired. After the holiday, there is a sharp dip in the treated series during the first month, but only slight additional reversal thereafter as the treated group is marginally below the control group for about a year and then converges with the control group in the later part of the sample. Taken at face value, this graph implies that housing stimulus gave a large boost to housing market activity during the policy with very weak reversal after the policy (apart from the short-term timing effect shown by the spike and dip right around the stimulus end date analysed in section 1.5.1 above).[22] However, we argue that this both overstates the positive impact of the stimulus policy and understates the slump after the end of the policy.

The issue with the analysis in Figure 11 is that treatment assignment (whether a transaction takes place in the £125,000–£175,000 price range) is endogenous to movements across bracket cutoffs. The stamp duty holiday creates an incentive to move into the treated price bracket from both sides. At the upper end of the range, the holiday creates a new notch at £175,000 that induces agents to move from a region above the cutoff to a point just below the cutoff (bunching). We have shown in section 1.4 above that bunching responses at £175,000 do indeed occur, and this increases activity in the treated range compared to the control range. At the lower end, the holiday eliminates the notch at £125,000 and

[21]As described in section 1.3.1, a stamp duty relief scheme was implemented for first-time buyers in the price range £125K–£250K between 25 March 2010 and 24 March 2012 (*after* the end of the stamp duty holiday). Since we are also interested in estimating reversal after the stamp duty holiday, it is important to make sure that the first-time buyers' relief scheme is not a confounding factor during the reversal period. This motivates using a control range (£175K–£225K) just above the treatment range (£125K–£175K), ensuring that both groups fall within the range eligible for first-time buyers' relief and therefore face the same incentive from this scheme. There could still be a concern that the treatment and control range respond differently to the first-time buyer incentive, which would be a confounding factor in the reversal estimates. To alleviate this concern, we drop all transactions claiming first-time buyers' relief throughout the analysis in this section. Including those observations only strengthens our findings below of incomplete reversal after the end of the stamp duty holiday.

[22]Note that the control group also features a (much smaller) spike and dip around the end of the stamp duty holiday driven by the Christmas/New Year effect as discussed in section 1.5.1 above.

therefore induces bunchers at this cutoff to move back into the hole above the cutoff. We have shown that the disappearance of bunching at £125,000 also occurs, and this further increases activity in the treated range compared to the control range. Hence, the positive effect of housing stimulus in Figure 11 combines the true effect on overall activity levels with endogenous price responses resulting from the change in the location of the notch.

There are two ways of dealing with this endogeneity issue. The simplest way is to widen the treatment range on each side (below £125,000 and above £175,000) in order to ensure that any price manipulation around notches occurs *within* the treatment range and so does not affect measured activity levels in this range. By including transactions outside the tax holiday area in the treatment group, this strategy captures an intent-to-treat effect and therefore understates the impact on the actually treated. We consider this intent-to-treat strategy in Appendix Figure 16, but here we focus instead on a more sophisticated way of dealing with endogeneity. This strategy exploits the fact that we have monthly bunching estimates of price responses to notches and can therefore directly control for it. That is, we may consider the number of transactions in different price brackets adjusted for the effect of bunching behavior in each month. To be precise, in every month, the estimated bunching mass just below £125,000 is reallocated to the treatment range £125,000–£175,000 while the estimated bunching mass just below £175,000 is reallocated to the control range £175,000–£225,000. By using these bunching-adjusted counts in our difference-in-differences strategy, we avoid bias from selection into treatment.

Figure 12 shows the results from this bunching-adjusted strategy. Panel A shows the normalized logs of the monthly number of transactions in the treatment and control ranges exactly as in Figure 11. It is visually clear that this strategy results in effects of housing stimulus that are qualititatively similar, but considerably smaller, and that there is a stronger lull in activity after the end of the stamp duty holiday. Panel A also suggests that the lull in activity lasts for approximately 12 months, after which the two series are completely parallel again. Panel B shows the cumulative sums of the two series in panel A as well as the cumulative sum of the differences between the two series (in green diamonds) in order to emphasize the effects we are studying. Panel B confirms that the two series track each other before the stimulus, diverge gradually during the stimulus period, and then converge for around 12 months until they revert to their pre-stimulus, parallel trends.

In order to quantify the effects of the stimulus, we run the following regression on a panel of monthly activity levels in price bins of £5,000 (over the range £125,000–£225,000) between September 2006 and October 2012

$$
\begin{aligned}
n_{it} &= \alpha_0 Pre_t + \alpha_H Hol_t + \alpha_R Rev_t + \alpha_P Post_t + \alpha_T Treated_i \\
&\quad + \beta_H Hol_t \times Treated_i + \beta_R Rev_t \times Treated_i + \beta_P Post_t \times Treated_i + \nu_{it}, \qquad (1.5.2)
\end{aligned}
$$

where n_{it} is the log number of transactions in price bin i and month t, Pre_t is a dummy for the pre-period September 2006–August 2008, Hol_t is a dummy for the stamp duty holiday period September 2008–December 2009, Rev_t is a dummy for the post-holiday reversal period January–December 2010, $Post_t$ is a dummy for the later months January 2011–October 2012, $Treated_i$ is a dummy for the treated price range £125,000–£175,000, and finally ν_{it} is an error term that we allow to be clustered at the monthly level.[23] The coefficients we are interested in are β_H (positive effect during stimulus) and β_R (negative effect after stimulus due to re-timing).

Panel A of Figure 12 shows our estimates of the coefficients β_H, β_R and β_P. The coefficient $\hat{\beta}_H = 0.20\ (0.022)$ implies that average monthly activity was approximately 20% higher during the holiday than it would have been in the absence of stimulus, corresponding to an extensive margin elasticity η as defined in section 1.2.3 of $20.62\ (2.18)$.[24] The coefficient $\hat{\beta}_R = -0.08\ (0.032)$ implies that average monthly activity was about 8% lower in the first year after the stimulus than it otherwise would have been. Together, these estimates imply that 31% of the additional activity created by the stimulus program was a timing response by people bringing forward their purchases in order to benefit from the tax cut, while the remaining 69% was a permanent, extensive margin effect.[25] This implies that the long-run elasticity is $14.3\ (3.26)$.

[23] Since we run the difference-in-differences regression (1.5.2) using bunching-adjusted activity levels in £5K bins, we have to reallocate bunching mass below the two cutoffs to specific £5K bins above the cutoffs. We reallocate bunching mass below a cutoff to the five bins above the cutoff in proportion to the amount of missing mass (difference between the estimated counterfactual mass and the observed mass) in each bin. Furthermore, since activity levels are adjusted using *estimated* bunching at the thresholds, we are introducing measurement error to our dependent variable coming from misspecification of the counterfactual when calculating the amount of bunching at £125K and £175K. However, since this measurement error is effectively noise in the dependent variable, it does not cause bias in our estimates, but simply increases our standard errors.

[24] The elasticity is estimated as $\hat{\eta} = \beta_H / [\Delta t / (1+t)]$

[25] The estimate of total reversal as a share of total stimulus is calculated as $-\left(12\hat{\beta}_R\right) / \left(16\hat{\beta}_H\right)$.

Since the end date of the reversal period (December 2010) was chosen visually as the point at which the two series become parallel again, there might be a concern that our estimate of total reversal is sensitive to the choice of this end date. In order to address this, Panel C of Figure 12 shows how this result changes as a different end date is chosen. The green diamonds show estimates of total reversal as a share of total stimulus as the regression (1.5.2) is performed using different reversal period cutoffs, and the grey shaded area depicts the 95% confidence interval around these estimates.[26] The reversal estimate is not sensitive to this choice, never rising above 40%, and we can always confidently reject the presence of full reversal.

When considering the simpler intent-to-treat strategy described above (see Figure 16), the effects are qualitatively similar but quantitatively somewhat weaker as one would expect. The intent-to-treat strategy produces larger reversal as a share of stimulus (40–50%) than the bunching-adjusted strategy, but we can still reject full reversal in all specifications.

These reversal findings stand in sharp contrast to Mian & Sufi (2012), who find complete and swift reversal following a short (1 month) stimulus program offering car transaction subsidies in the US. The contrast between our findings and those of Mian & Sufi (2012) may suggest that stimulus policies that are of extremely short duration, such as the one they study, do not give households sufficient time to respond along the extensive margin and therefore have only short-term timing effects. Hence, our findings highlight the importance of the length of the stimulus program. Of course, while the strength of reversal is important for evaluating stimulus, it does not by itself indict or validate such policies as their key rationale is to create more economic activity when the economy is slack (even if this comes at the expense of less economic activity when the economy is tight). The next section provides a rough estimation of the immediate increase in real economic activity created by the UK housing stimulus program.

It is important to note that our quasi-experimental micro approach to evaluating stimulus policy does not capture potential general equilibrium or multiplier effects. If the program had a salutary effect on the housing market and macroeconomy as a whole, this effect would be present in both treatment and control groups and therefore not show up in our difference-in-differences estimates. Besides general equilibrium and multiplier

[26] The point estimates are calculated as $-\left(\sum_t Rev_t \times \hat{\beta}_R\right) / \left(16\hat{\beta}_H\right)$, where $\sum_t Rev_t$ denotes the length of the reversal period in the particular regression. Standard errors are computed by the delta method.

effects, a source of spillovers between treatments and controls may arise from *real estate chains*, i.e. linked house transactions whereby someone selling a house in the treatment range is simultaneously buying a house in the control range. Bias from chain effects can be reduced or eliminated by considering control ranges further away from the treatment range, but such strategies create other problems with comparability and parallel trends. The key thing to realize is that potential chain effects unambiguously work against us and create attenuation bias, and so the (large) stimulus estimates we obtain by comparing neighboring price ranges are, if anything, conservative.

1.5.3. GDP Effects of Stimulus. While we have established that the stamp duty holiday had a large effect on transaction volume in the housing market (and therefore on household mobility), a motivation for the policy was also to stimulate real economic activity through larger household spending driven by the complementarities between moving house and spending. Investigating the spending effect of the UK housing stimulus program also allows for a comparison between our findings and previous work on the consumer spending effect of fiscal stimulus such as income tax rebates (e.g. Shapiro & Slemrod 2003a,b; Johnson *et al.* 2006; Agarwal *et al.* 2007; Kreiner *et al.* 2012). A fully rigorous analysis of the effects of housing transactions on expenditure is beyond the scope of this paper, but we perform some back-of-the-envelope calculations to shed light on the likely magnitude of these effects.

Using data from the UK Living Costs and Food Survey, we estimate in Appendix table 4 that households spend roughly an additional 1.6% of the value of their home on repairs, improvements, furnishings, appliances and other durable goods when they move. This is a conservative estimate compared to similar calculations for the US (Siniavskaia 2008; Zillow.com 2012). Estate agents' fees average 1.98% of the value of the house and other commissions come to 1.24%, giving an estimate of the total expenditure accompanying a house transaction of 4.8% of the house value. Denoting this estimate by ϕ, the immediate impact of the policy on GDP is $\Delta GDP = \phi h_v^m \Delta n$ where h_v^m is the average value of houses bought during the stimulus, and Δn is the number of additional transactions resulting from the policy. To arrive at an estimate of the effectiveness of the policy that is comparable to other stimulus policies, we scale it by the foregone tax revenue, $\Delta Tax = \tau_0 h_v^m n_0$ where $\tau_0 = 1\%$ is the pre-stimulus tax rate, and n_0 is the counterfactual number of transactions in the price range affected by the stimulus. In the previous section, we estimated $\Delta n/n_0$ to be $\beta_H = 0.20$, and so we arrive at an estimate of the effect on economic activity per dollar of tax cut equal to $\Delta GDP/\Delta Tax = \phi\beta_H = 0.96$.[27]

These calculations suggest that the stamp duty holiday was not only successful in stimulating housing market activity, but also provided a significant boost to real economic activity through the strong complementarities between moving house and consumer spending. These rough calculations exclude other indirect effects, for example labor market effects of increased mobility and Keynesian multiplier effects. As a benchmark, the previous work cited above on fiscal stimulus through income tax rebates found significantly smaller effects on consumer spending (0.2–0.7 dollars of spending per dollar of tax cut, as opposed to about 1 dollar of spending here). Overall, our findings suggest that transaction tax cuts (or subsidies) can be very effective at stimulating both housing market activity and real economic activity during downturns.

1.5.4. Extensive Margin Responses to a Permanent Tax Cut. On 16 March 2005, the bottom notch was moved from £60,000 to £120,000. The reform took effect immediately after its announcement, and while a reform of this kind had been expected, the exact timing and details were not. Since this was a permanent reform, studying its impact over an extended period after its implementation will allow us to analyze the extensive margin effects of permanent reforms (since potential timing effects will only affect the months just after the reform). It is also worth noting that this reform was implemented during the height of the housing market boom, in sharp contrast to the stamp duty holiday implemented at the bottom of the recession.

The reform cut the tax from 1% to 0% over the price range £60,000 to £120,000 while leaving the tax unchanged in neighbouring price ranges, which again presents us with the opportunity to pursue a difference-in-differences strategy. The issue that treatment assignment is endogenous to price responses to the movement of the notch is present in exactly the same way as for the stamp duty holiday, and so we address it in the same way by using monthly bunching estimates to account for price responses. Figure 13 shows the results from our bunching-adjusted difference-in-differences strategy. Panel A shows the

[27] Appendix Table 4 shows details of the calculations and their sensitivity to using the intent-to-treat estimate of β_H discussed in the previous subsection as well as an alternate estimate of households' additional expenditure.

normalized log counts of monthly transaction volumes in the treatment range £60,000–£120,000 (blue circles) and the control range £120,000–£180,000 (orange crosses) together with the estimated treatment effect from a regression analogous to equation (1.5.2), while panel B shows the cumulative sums of the normalized log counts in the treatment and control ranges. As panel A shows, the treatment and control ranges were parallel in the months leading up to the reform, and then diverged sharply immediately following the reform. The estimated coefficient $\hat{\beta}_P = 0.23\ (0.018)$ implies that this permanent reform increased monthly transaction volumes by approximately 23% on average, implying an extensive margin elasticity η of $23.2\ (1.86)$. This effect is considerably larger than the permanent effect of the stamp duty holiday stimulus, consistent with the idea that the permanent effect of tax changes is increasing in the length of the time horizon of the policy as discussed in section 1.5.2.

1.5.5. Long Run Revenue and Welfare Impacts of the Transaction Tax. With our estimates of the intensive margin house price elasticity ε and the long-term extensive margin elasticity η, we can evaluate the long-run revenue and welfare impacts of the transaction tax as outlined in section 1.2.3. Table 3 presents the results from applying equation (1.2.14) for the marginal cost of funds (MCF) separately in each bracket, and applying equation (1.2.15) for the Laffer rate. For the extensive margin elasticity η, we use two possible values: the long-term extensive margin elasticity of 14.3 estimated from the stamp duty holiday in section 1.5.2, and the long-term elasticity of 23.2 estimated from the permanent reform in section 1.5.4. For the intensive margin house price elasticity ε, we use the estimates obtained from bunching at notches shown in Table 2. The calculation of MCF in each bracket is based on the intensive elasticity obtained from the notch at the bottom of the bracket, while the calculation of the Laffer rate is based on the intensive elasticity obtained from the 1% bracket (as this price bracket contains the average house price in the UK).

The MCF increases strongly as we move up the tax schedule and is extremely large in the higher brackets. The MCF in the first tax bracket is 1.1–1.3, while the MCF in the top tax bracket is about 18 under the smaller extensive margin elasticity, or beyond the Laffer rate (negative MCF) under the larger extensive margin elasticity. These MCF estimates are very large compared to standard tax instruments. For example, Kleven & Kreiner (2006) estimate an MCF for the UK personal income tax that ranges between 1.1 for the bottom earnings decile and 2.68 for the top earnings decile. Similarly, Dahlby (2008) reports that the overall MCF for the UK income tax is about 1.2, and that for the US income tax it is about 1.1. The strong responsiveness to the transaction tax also implies that the Laffer rate (Pareto bound) is very small, between 4 and 7%, as compared to 70–80% for the income tax in the US (Saez *et al.* 2012). Overall, the estimates in Table 3 raise important questions about the desirability of transaction taxes as a long-run policy instrument.

1.6. Conclusion

This paper has studied the impact of property transaction taxes on the housing market, using unique administrative data on every property transaction in the UK from 2004–2012 and compelling quasi-experimental variation created by notches, tax reforms, and stimulus. We have presented evidence on the effects of transaction taxes on house prices as well as on the timing and volume of house purchases, including an analysis of the dynamics of adjustment to both anticipated and unanticipated tax changes. Using a variety of methods, we find that prices and especially activity levels in the housing market respond very strongly and quickly to transaction taxes. Our estimates imply that the marginal cost of funds for the property transaction tax is orders of magnitude larger than for other tax instruments, and that the revenue-maximizing tax rate (Laffer rate) is as low as 4–7%. Transaction taxes beyond this modest level are Pareto inefficient.

It should be noted that it is not the existence of notches *per se* that makes the transaction tax so distortionary. The large distortions arise from the strong underlying responsiveness to the tax, which we are able to identify using notches and other sources of exogenous variation. Moreover, our study of transaction taxes in the property market could also have implications for the potential effects of transaction taxes in other asset markets, including the transaction taxes on financial assets that have been discussed widely in recent years.

Our findings from the 2008–2009 stamp duty holiday contribute to the scant micro evidence on the effectiveness of fiscal stimulus and, in particular, present some of the first evidence on the effectiveness of using temporary tax changes to stimulate the housing market during economic downturns. The 16-month stamp duty holiday was enormously successful in stimulating housing market activity, increasing the volume of house transactions by as much as 20% in the short run (due to timing and extensive responses) followed by a smaller slump in activity after the policy is withdrawn (as the timing effect is cancelled out). Due to the complementarities between moving house and consumer spending, these stimulus effects translate into GDP effects that are considerably larger than what has been found for other forms of fiscal stimulus such as income tax rebates. Thus, the stimulus was successful both at stimulating activity in the housing market, and at stimulating the real economy. More generally, these findings suggest that policies that reduce the cost of housing transactions are likely to be effective as stimulus policies. Beyond the UK, this finding lends support to the reasoning behind the homebuyer tax credit introduced by the 2009 Stimulus Bill in the US.

An interesting dynamic question remains regarding the ability of asset transaction taxes to affect the emergence of asset-price bubbles and the cyclicality of the economy more generally. Addressing this issue raises some daunting empirical challenges, ideally requiring exogenous variation in transaction taxes across economies, and so is left for future research.

TABLE 1. RESIDENTIAL PROPERTY TAX NOTCHES

<table>
<tr><th>Date Range
Price Range</th><th>1 Dec 2003
to
16 Mar 2005</th><th>17 Mar 2005
to
22 Mar 2006</th><th>23 Mar 2006
to
2 Sep 2008</th><th>3 Sep 2008
to
31 Dec 2009</th><th>1 Jan 2010
to
5 Apr 2011</th><th>6 Apr 2011
to
21 Mar 2012</th><th>22 Mar 2012
to
April 2013</th></tr>
<tr><td>0 - £60K</td><td>0</td><td rowspan="2">0</td><td rowspan="3">0</td><td rowspan="4">0</td><td rowspan="3">0</td><td rowspan="3">0</td><td rowspan="3">0</td></tr>
<tr><td>£60K - £120K</td><td rowspan="4">1</td></tr>
<tr><td>£120K - £125K</td><td rowspan="3">1</td></tr>
<tr><td>£125K - £175K</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td></tr>
<tr><td>£175K - £250K</td><td>1</td></tr>
<tr><td>£250K - £500K</td><td>3</td><td>3</td><td>3</td><td>3</td><td>3</td><td>3</td><td>3</td></tr>
<tr><td>£500K - £1000K</td><td rowspan="3">4</td><td rowspan="3">4</td><td rowspan="3">4</td><td rowspan="3">4</td><td rowspan="3">4</td><td>4</td><td>4</td></tr>
<tr><td>£1000K - £2000K</td><td rowspan="2">5</td><td>5</td></tr>
<tr><td>£2000K - ∞</td><td>7</td></tr>
</table>

Notes: The table shows how the stamp duty land tax schedule for residential property has varied over time. Each column represents a time period during which the tax schedule was constant. The rows represent price ranges, and the entry in each cell is the tax rate that applies to that price range in the time period.

TABLE 2. INTENSIVE ELASTICITY ESTIMATES

Notch Point (£000s)	Period	Δt (%-points)	Δh_v	Intensive Elasticity ε Structural	Reduced Form
60	11/2004 – 3/2005	1	3,500 (1,026.1)	0.17 (0.097)	0.34 (0.200)
120	4/2005 – 3/2006	1	5,000 (282.2)	0.09 (0.038)	0.17 (0.020)
125	3/2006 – 9/2008	1	5,000 (534.0)	0.09 (0.052)	0.16 (0.034)
175	9/2008 – 12/2009	1	5,000 (1,981.8)	0.05 (0.051)	0.08 (0.065)
125	1/2010 – 10/2012	1	5,000 (274.6)	0.09 (0.027)	0.16 (0.018)
250	11/2004 – 10/2012	2	10,000 (1,997.0)	0.05 (0.025)	0.08 (0.032)
500	11/2004 – 10/2012	1	10,000 (3,808.7)	0.03 (0.024)	0.04 (0.031)

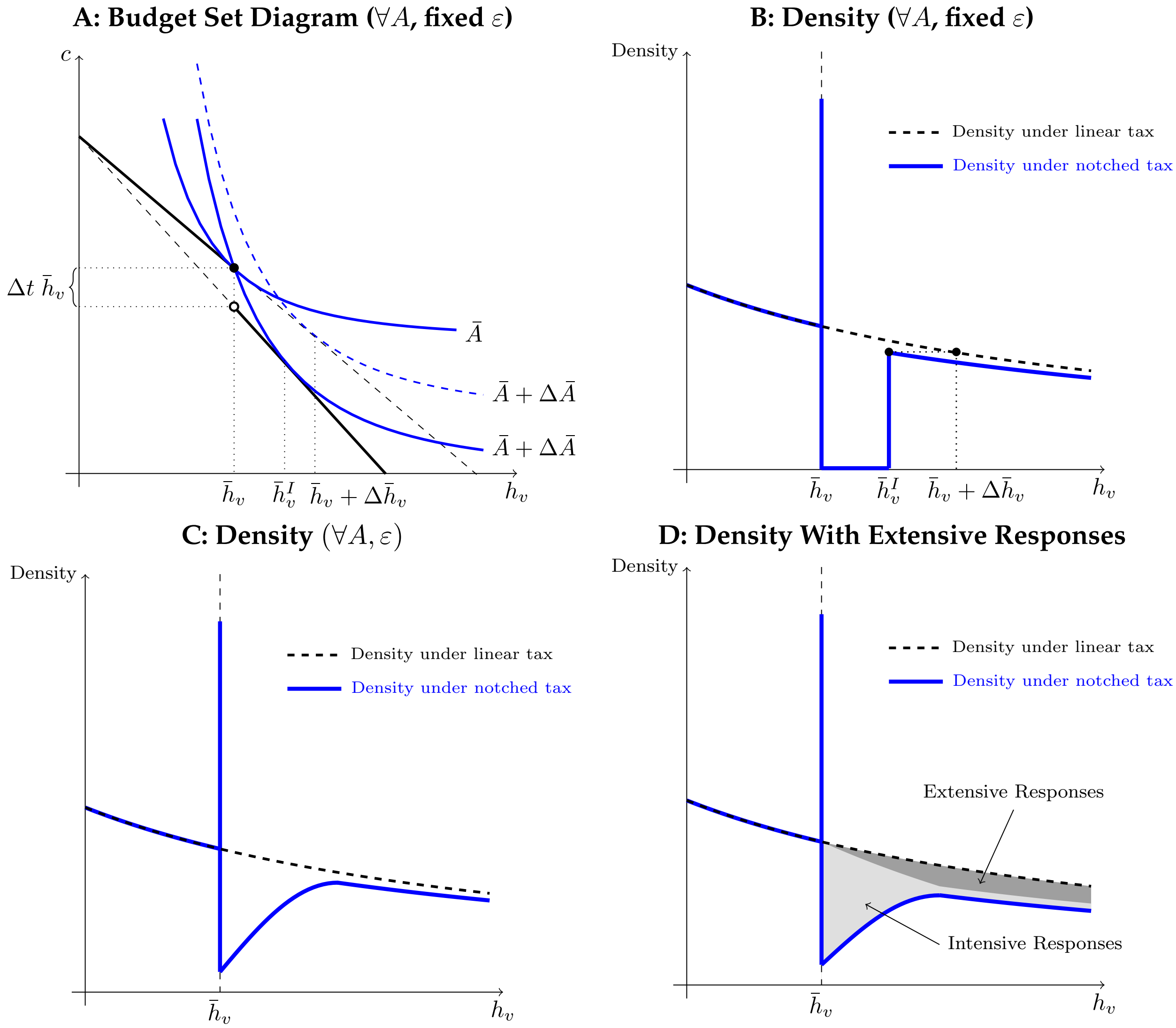
Notes: The table shows the absolute values of the structural and reduced-form estimates of the house price elasticity implied by bunching at the notches. The *From* and *To* columns demarcate the period the notch was in place. The Δt column is the percentage-point jump in the tax rate at the notch. The $\Delta \bar{h}_v$ column shows our estimate of the average house price change implied by our estimated bunching mass using equation (1.2.7), and it's bootstrapped standard error in parentheses. The final two columns show our estimate of the structural elasticity implicitly defined by equation (1.2.9) with its standard error (derived by the delta method) in parentheses and our estimate of the reduced-form elasticity using equation (1.2.10) with its standard error (derived by the delta method) in parentheses.

TABLE 3. MARGINAL COST OF FUNDS AND LAFFER RATES FOR A TRANSACTION TAX

Bracket (£000s)	**Rate (%)**	**Intensive Elasticity** $\bar{\varepsilon}$	**Marginal Cost of Funds** $\bar{\eta} = 14.3$	$\bar{\eta} = 23.2$
125 – 250	1	0.16	1.17	1.31
250 – 500	3	0.08	1.72	3.11
500 – 1,000	4	0.04	2.23	9.42
1,000 – 2,000	5	0.09	3.18	L
2,000 –	7	0.13	17.86	L
Laffer Rate t_L		0.16	7.43	4.47

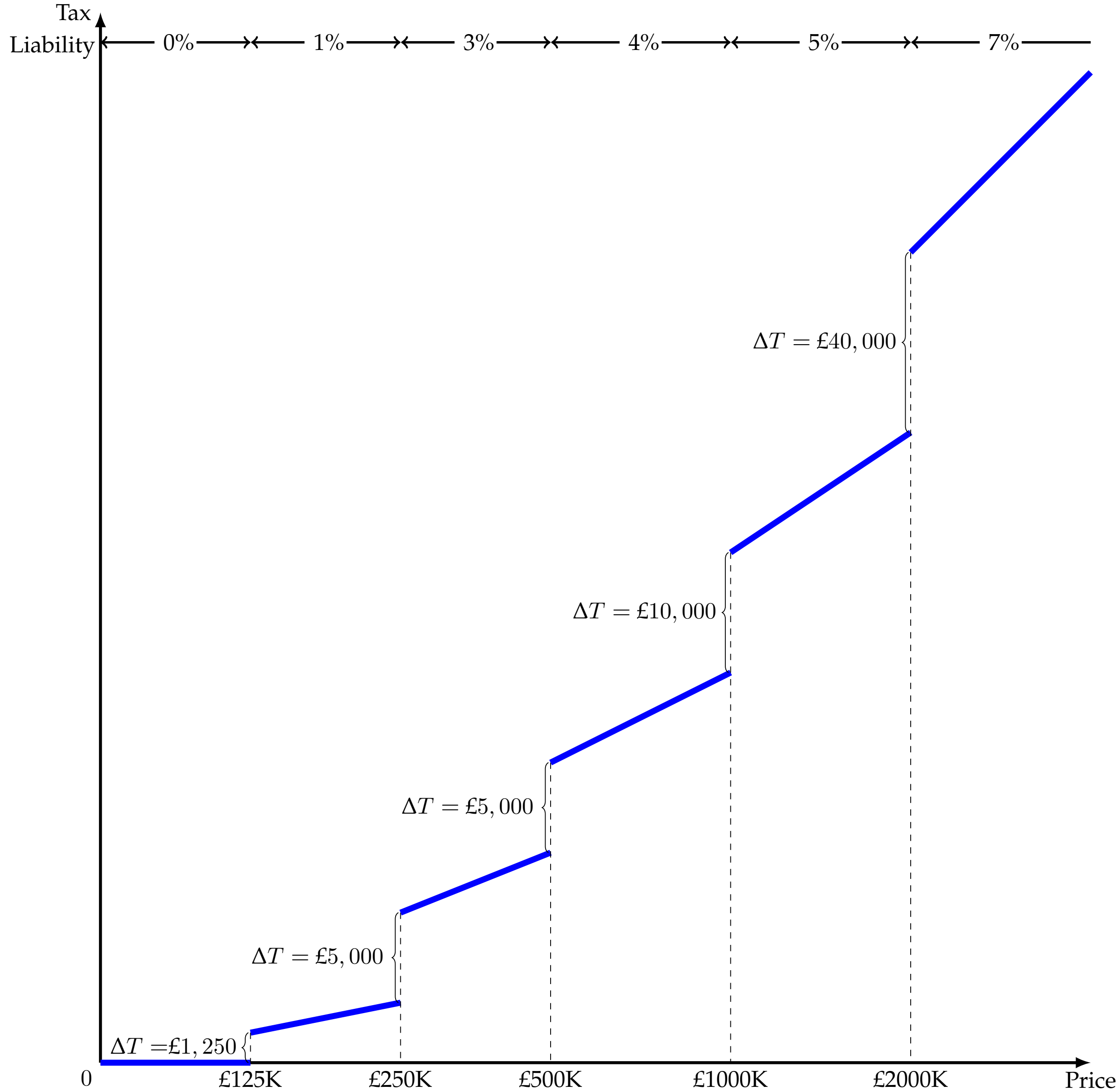
Notes: The table shows the marginal cost of funds in the different brackets of the stamp duty schedule in place as of 2013, and the Laffer rate for the stamp duty. The marginal cost of funds is given by equation (1.2.14). The Laffer rate is given by equation (1.2.15). A value L indicates the rate is beyond the Laffer rate. The $\bar{\varepsilon}$ column shows the intensive margin elasticity used for the calculations. For the marginal cost of funds, this is the absolute value of the reduced-form intensive margin elasticity estimated from the notch at the bottom of the bracket, while the the value from the 1% bracket (containing the average house value) is used for the Laffer rate calculation. $\bar{\eta}$ is the absolute value of the extensive margin elasticity.

FIGURE 1. THEORETICAL FIGURES



Notes: Figure 1 illustrates the implications of a notched transaction tax schedule in a budget set diagram (Panel A) and density distribution diagrams (Panels B-D). The budget set diagram in panel A (depicting preferences as in equation (1.2.1) and the budget set given by equation (1.2.2) in (h_v, c)-space) illustrates intensive responses among individuals with heterogeneous housing preferences A, but a specific demand elasticity ε. The notch creates bunching at the cutoff $\overline{h}_v$ by all individuals in a preference range $(\overline{A}, \overline{A} + \Delta\overline{A})$, who would have bought houses on the segment $(\overline{h}_v, \overline{h}_v + \Delta\overline{h}_v)$ in the absence of the notch. The marginal bunching individual at $\overline{A} + \Delta\overline{A}$ is indifferent between the notch point $\overline{h}_v$ and the best interior location $\bar{h}_v^I$. No individual is willing to locate between $\overline{h}_v$ and $\bar{h}_v^I$, and hence this range is completely empty. The density of property values corresponding to the budget set diagram (all A, one specific ε) is shown in Panel B. Since the behavioral response in Panels A-B depends on the size of the demand elasticity ε (and converges to zero for completely price inelastic buyers), the density in the full population (all A, ε) can be illustrated as in Panel C where some individuals are willing to buy just above the notch point. In addition to intensive responses, the notch creates extensive responses above the cutoff by individuals close to the indifference point between buying and not buying ($q \approx q^*$, where q^* is defined in equation (1.2.4)). However, such extensive responses will be negligible *just* above the cutoff. Intuitively, if an individual prefers buying a house slightly above $\overline{h}_v$ in the absence of the notch, then he will be better off by buying a house at $\overline{h}_v$ (which is almost as good) than not buying at all in the presence of the notch. This reasoning implies that extensive responses affect the density as illustrated in Panel D.

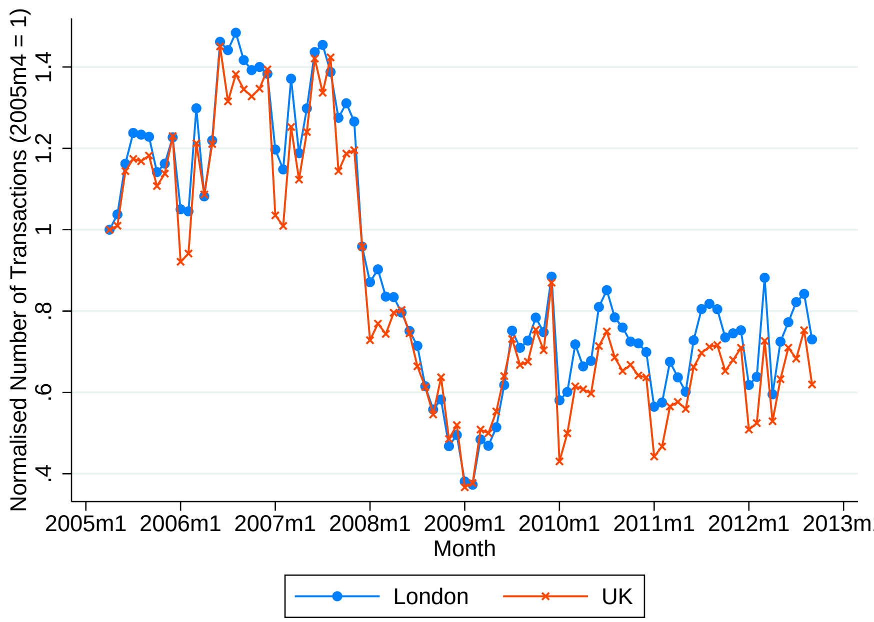
FIGURE 2. TAX SCHEDULE IN MARCH 2013



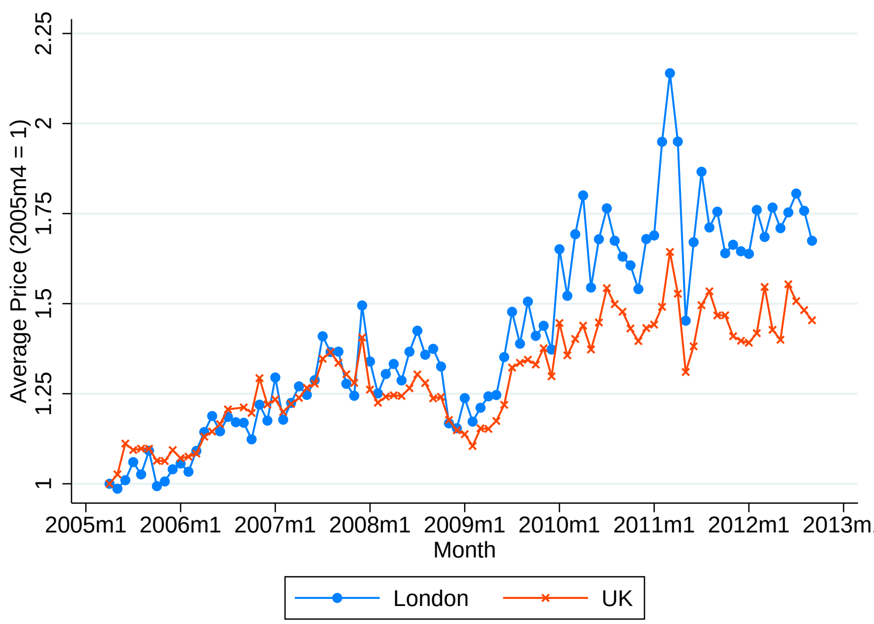
Notes: Figure 2 shows the stamp duty land tax schedule for residential properties in place in March 2013 graphically as the solid blue line. The tax liability jumps discretely at the notches at £125,000, £250,000, £500,000, £1,000,000 and £2,000,000. Within the brackets defined by these notches, the tax rate is constant, and applied to the whole transaction price at the rates shown along the top of the figure.

FIGURE 3. SUMMARY STATISTICS

A: Number of Transactions

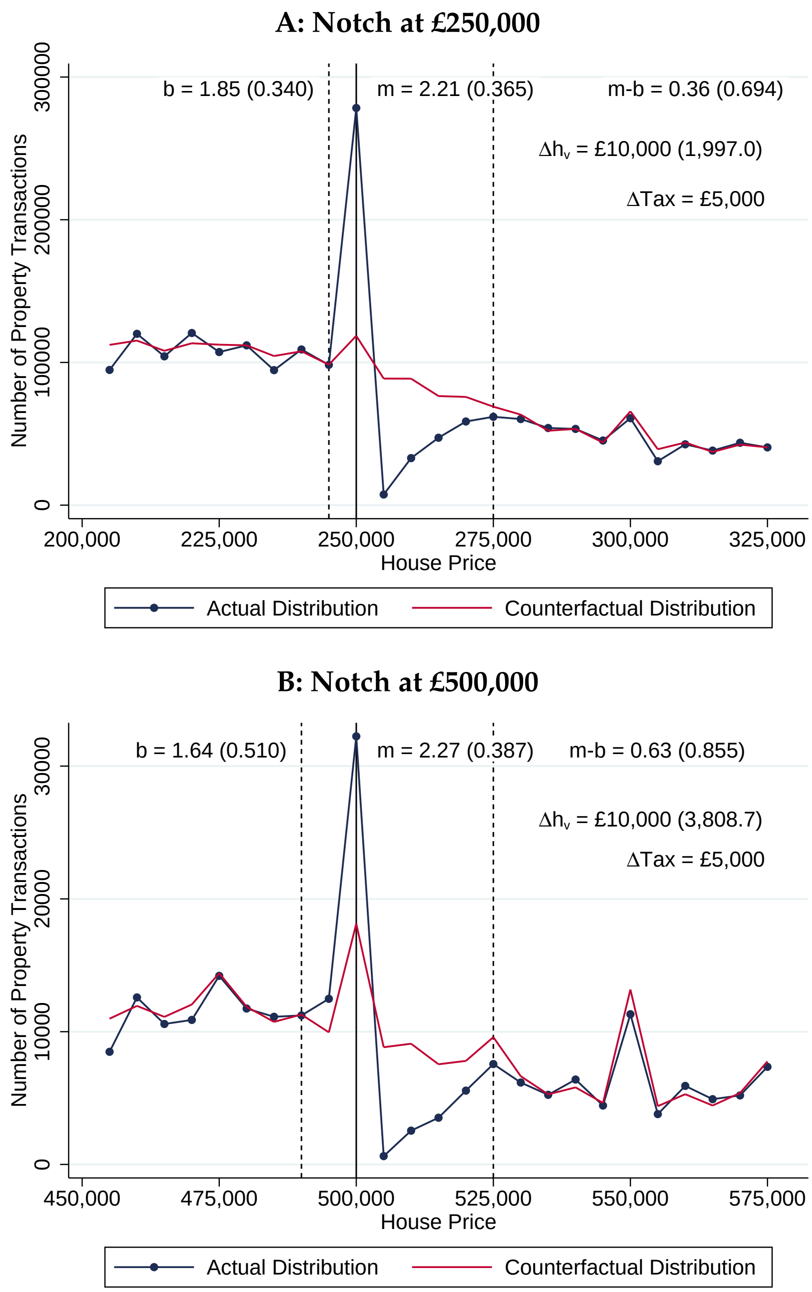


B: Average Price



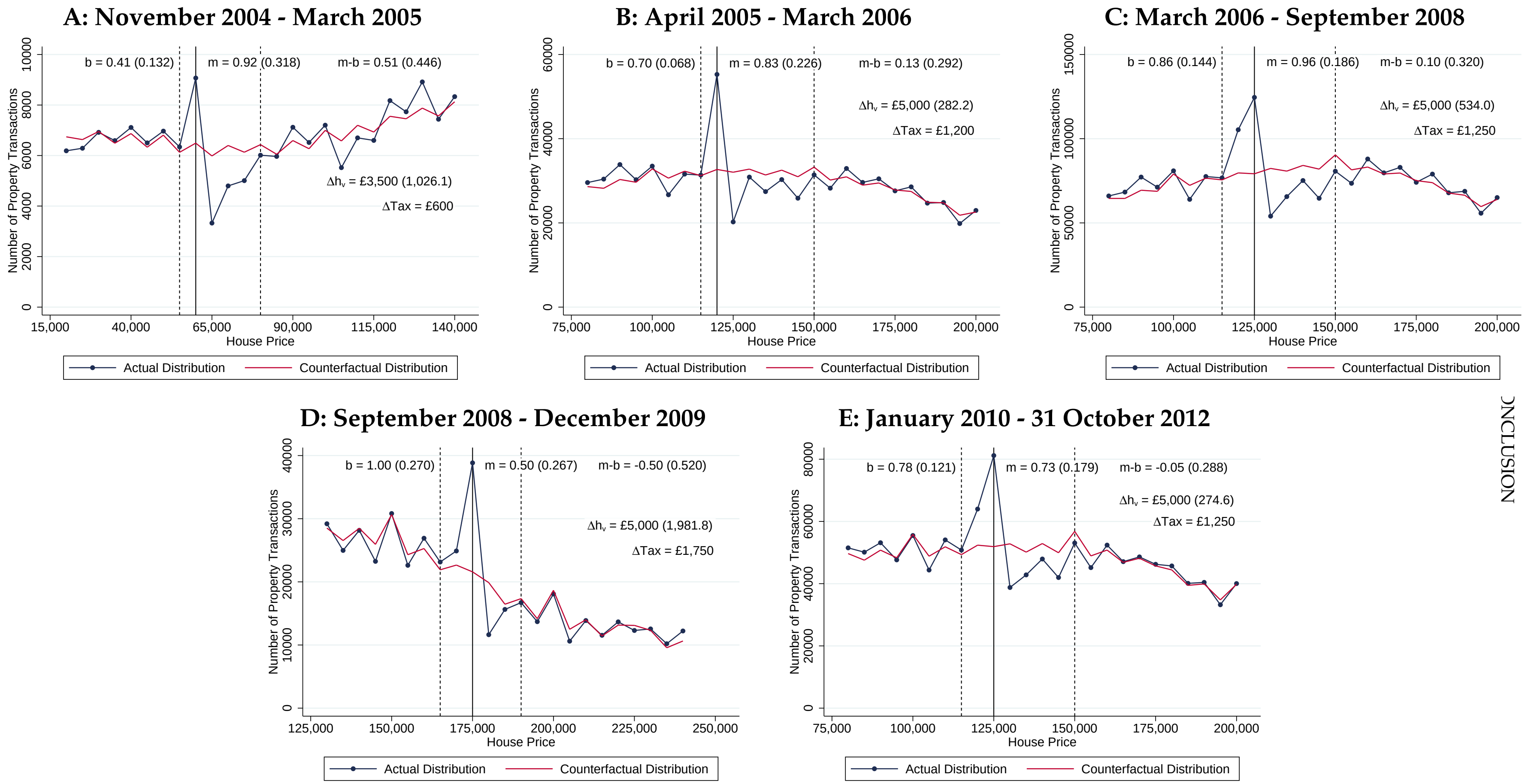
Notes: Panel A shows the monthly average price of property transactions relative to the average price in April 2005 in London (blue circles) and the U.K. (orange crosses). The average price of property transactions in London during the period April 2005 - October 2012 was £345,360 and the average price in the U.K. during our data period was £199,479. Panel B shows the monthly total number of property transactions relative to the number that took place in April 2005 in London (blue circles) and the U.K. (orange crosses). The average monthly number of property transactions in London during the period April 2005 - October 2012 was 12,955 while the average monthly number of property transactions in this period in the U.K. was 103,561.

FIGURE 4. BUNCHING AND HOLES AROUND THE NOTCHES THAT REMAIN CONSTANT



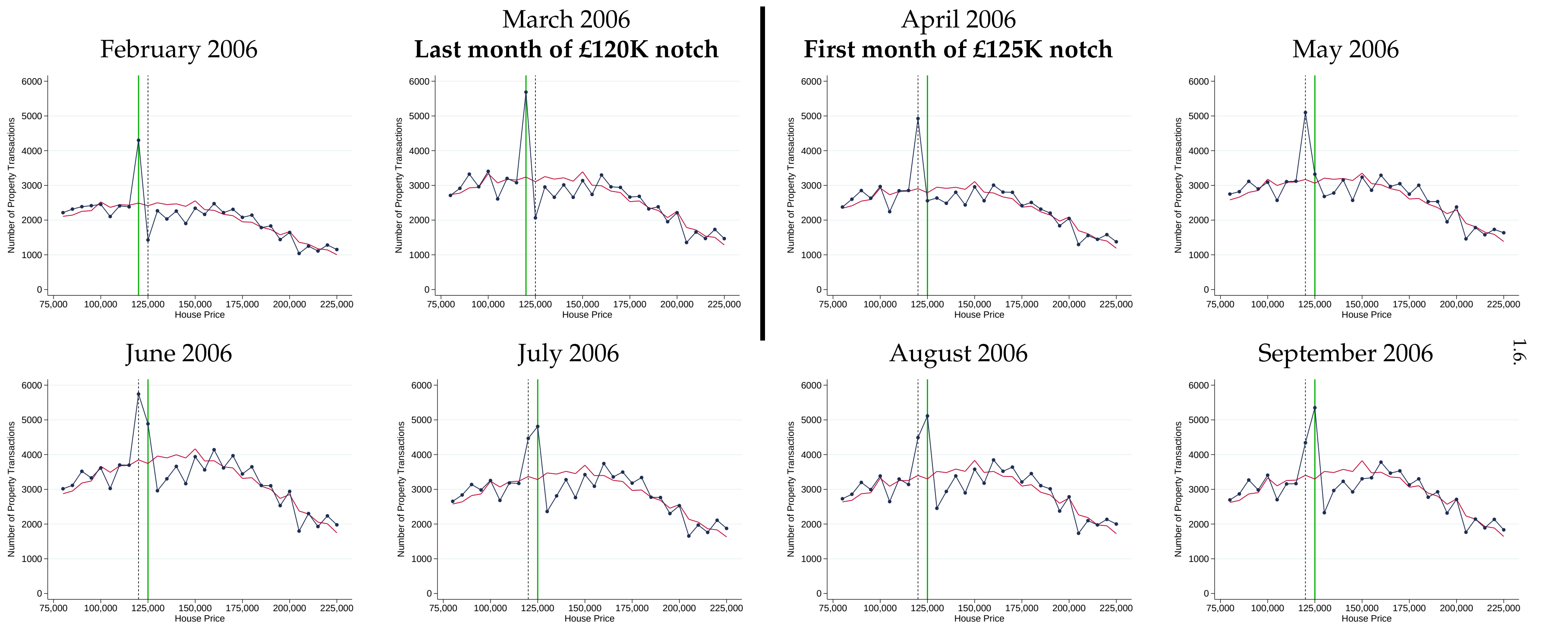
Notes: The figure shows the observed density of property transactions (blue dots) and our estimated counterfactual density (red line) around the notch at £250,000 where the tax liability jumps by £5,000 (from 1% to 3% of the transaction price) in panel A and around the notch at £500,000 where the tax liability jumps by £5,000 again (from 3% to 4% of the transaction price). The data used for these estimates excludes transactions that claim relief from the stamp duty land tax (except for those claiming first-time buyers' relief) as the regular tax schedule does not apply to these transactions. The counterfactual density is estimated as in equation (1.4.1), using bins £100 pounds wide and a polynomial of order 5. The vertical dashed lines denote the upper and lower bounds of the excluded region around the notch. The upper bound of the excluded region is chosen as the point where the observed density changes slope from positive to negative. The estimate of equation (1.4.1) controls for round number bunching at multiples of £500, £1,000, £5,000, £10,000, £25,000 and £50,000. Both the empirical and the counterfactual density are shown aggregated up to bins £5,000 wide. b is our estimate of the excess mass just below the notch scaled by the average counterfactual frequency in the excluded range, with its standard error shown in parentheses. m is our estimate of the missing mass above the notch scaled by the average counterfactual frequency in the excluded range, with its standard error shown in parentheses. $m - b$ is our estimate of the difference between the missing mass and the bunching mass, again with its standard error in parentheses. The figures also show the average house value change created by the notch, and the tax liability change at the notch. All standard errors are obtained by bootstrapping the procedure 200 times.

FIGURE 5. BUNCHING AND HOLES AROUND THE LOWER NOTCH



Notes: The figure shows the observed density of property transactions (blue dots) and our estimated counterfactual density (red line) around the lower notch in the residential property tax schedule where the tax liability jumps from 0 to 1% of the transaction price. Panel A shows the period 1 November 2004 to 16 March 2005 when the notch was at £60,000. Panel B shows the period 17 March 2005 to 22 March 2006 when the notch was at £120,000. Panel C shows the period 23 March 2006 to 2 September 2008 when the notch was at £125,000. Panel D shows the period 3 September 2008 to 31 December 2009 when the notch was at £175,000. Panel E shows the period 1 January 2009 to 31 October 2012 when the notch was at £125,000. The data used for these estimates excludes transactions that claim relief from the stamp duty land tax (excepting those who claimed first time buyers' relief) as the regular tax schedule does not apply to these transactions. The counterfactual density is estimated as in equation (1.4.1), using bins £100 pounds wide and a polynomial of order 5 in panels A, C, D and E and of order 4 in panel B. The vertical dashed lines denote the upper and lower bounds of the excluded region around the notch. The upper bound of the excluded region is chosen as the point where the observed density stops increasing and becomes decreasing (apart from spikes at round numbers). The estimate of equation (1.4.1) controls for round number bunching at multiples £500, £1,000, £5,000, £10,000, £25,000 and £50,000. Both the empirical and the counterfactual density are shown aggregated up to bins £5,000 wide. b is our estimate of the excess mass just below the notch scaled by the counterfactual density at the notch, with its standard error shown in parentheses. m is our estimate of the missing mass above the notch scaled by the counterfactual density at the notch, with its standard error shown in parentheses. $m - b$ is our estimate of the difference between the

FIGURE 6. DYNAMICS OF BUNCHING AT BOTTOM NOTCH AROUND MARCH 2006



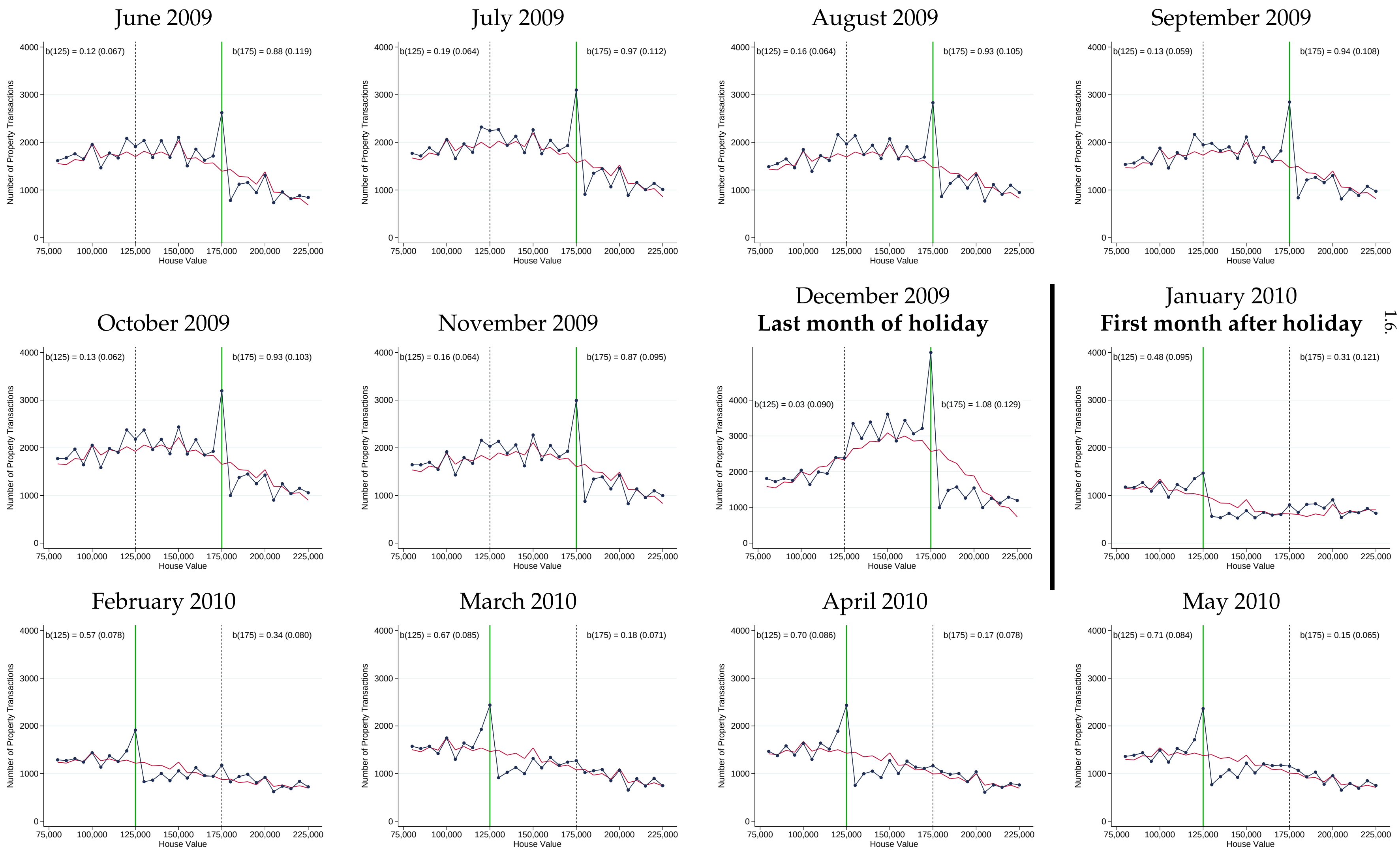
Notes: The figure shows the observed density of property transactions (blue dots) and our estimated counterfactual density (red line) in the region £75,000 – £225,000 separately for each month. On 23 March 2006, the bottom notch moved from £120,000 to £125,000. The estimation of the counterfactual is as described in section 1.4.1 and in the notes to figures 4 & 5. The estimation excludes data in the regions £115,000 – 140,000 and £170,000 – £190,000 and uses a polynomial of order 5. Animated versions of these figures that show the dynamics more vividly can be found at
http://personal.lse.ac.uk/kleven/Downloads/MyPapers/WorkingPapers/best-kleven_landnotches_april2013_videos.pdf

FIGURE 7. DYNAMICS OF BUNCHING AROUND THE BEGINNING OF STAMP DUTY HOLIDAY

1.6.

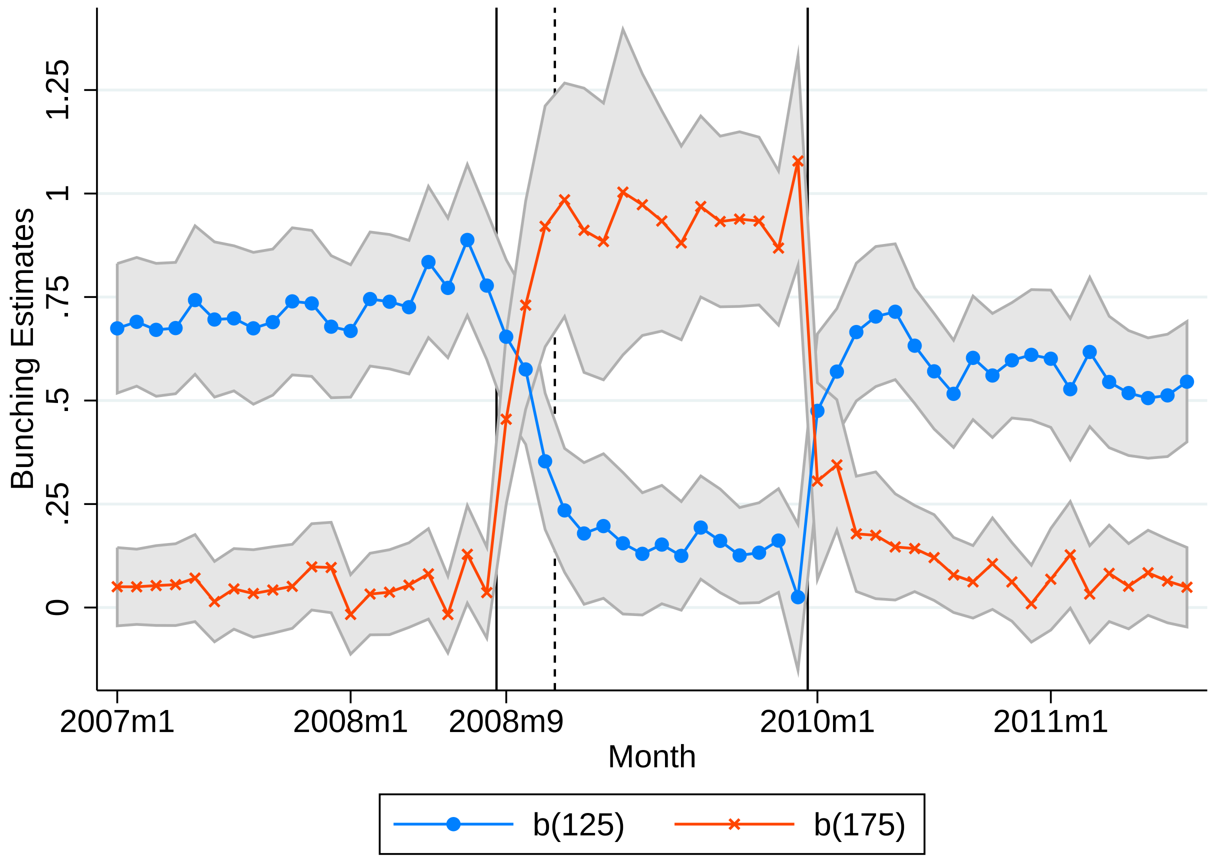
Notes: The figure shows the observed density of property transactions (blue dots) and our estimated counterfactual density (red line) in the region £75,000 – £225,000 separately for each month. On 3 September 2008, the bottom notch was moved unexpectedly from £125,000 to £175,000. The estimation of the counterfactual is as described in section 1.4.1 and in the notes to figures 4 & 5. The estimation excludes data in the regions £115,000 – 140,000 and £170,000 – £190,000 and uses a polynomial of order 5. Animated versions of these figures that show the dynamics more vividly can be found at
http://personal.lse.ac.uk/kleven/Downloads/MyPapers/WorkingPapers/best-kleven_landnotches_april2013_videos.pdf

FIGURE 8. DYNAMICS OF BUNCHING AROUND THE END OF STAMP DUTY HOLIDAY



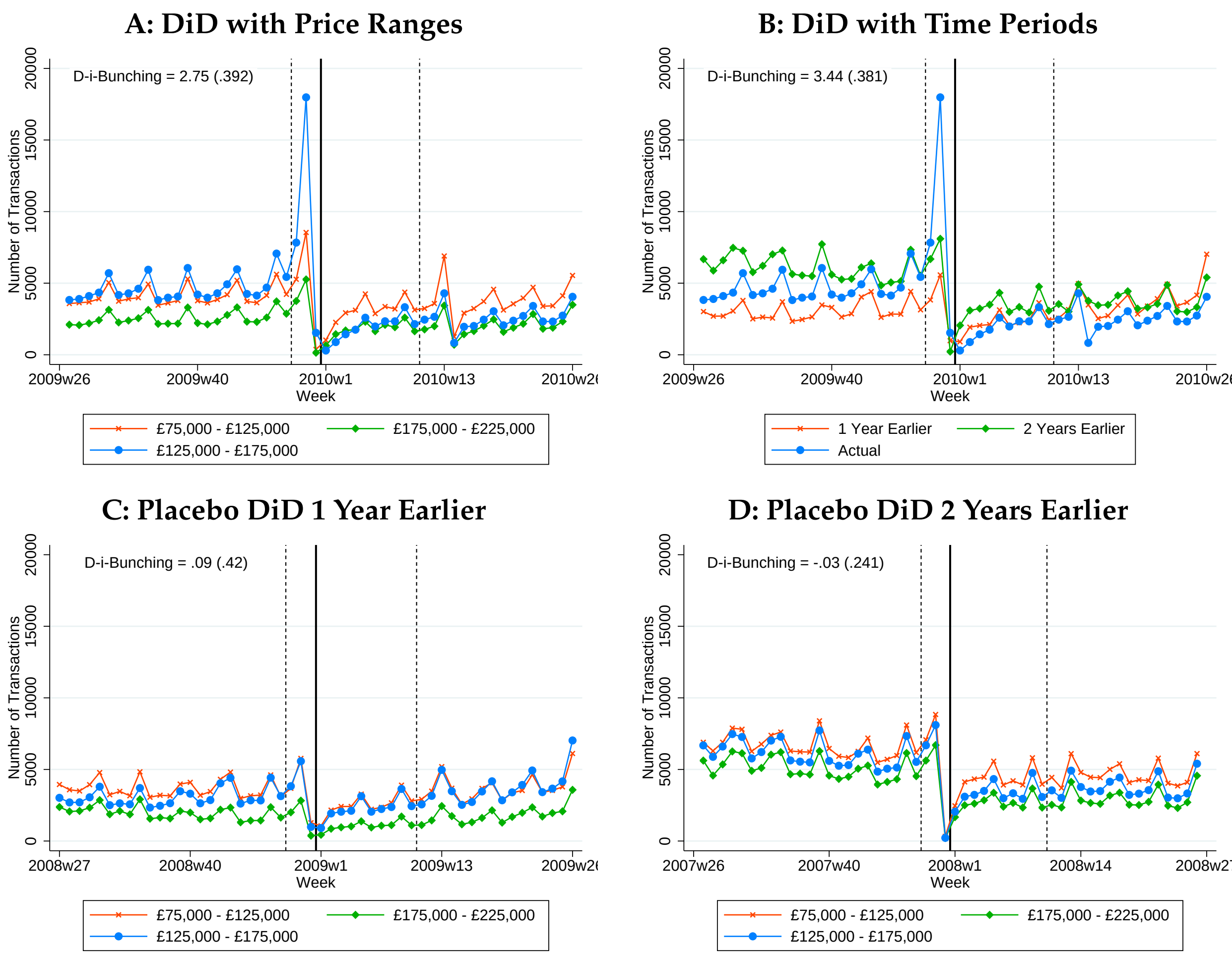
Notes: The figure shows the observed density of property transactions (blue dots) and our estimated counterfactual density (red line) in the region £75,000 – £225,000 separately for each month. On 1 January 2010, the bottom notch was moved back from £175,000 to £125,000 as announced previously. The estimation of the counterfactual is as described in section 1.4.1 and in the notes to figures 4 & 5. The estimation excludes data in the regions £115,000 – 140,000 and £170,000 – £190,000 and uses a polynomial of order 5. Animated versions of these figures that show the dynamics more vividly can be found at http://personal.lse.ac.uk/kleven/Downloads/MyPapers/WorkingPapers/best-kleven_landnotches_april2013_videos.pdf

FIGURE 9. BUNCHING ESTIMATES OVER TIME



Notes: The figure shows our estimates of $b\left(\bar{h}_v\right)$, the bunching mass just below $\bar{h}_v$ scaled by the counterfactual frequency at $\bar{h}_v$, by month from January 2007 to February 2011 and for two values of $\bar{h}_v$, £125,000 (blue circles) and £175,000 (orange crosses). The first vertical line is at September 2008 when the stamp duty holiday was unexpectedly announced, moving the notch from £125,000 to £175,000. The dashed vertical line is at December 2008 to represent the observation that house transactions take up to 90 days to conclude, and so some inertia in the bunching responses is to be expected. The second vertical line is at December 2009 when the stamp duty holiday came to an end as anticipated, and the notch was moved from £175,000 back down to £125,000.

FIGURE 10. TIME NOTCH

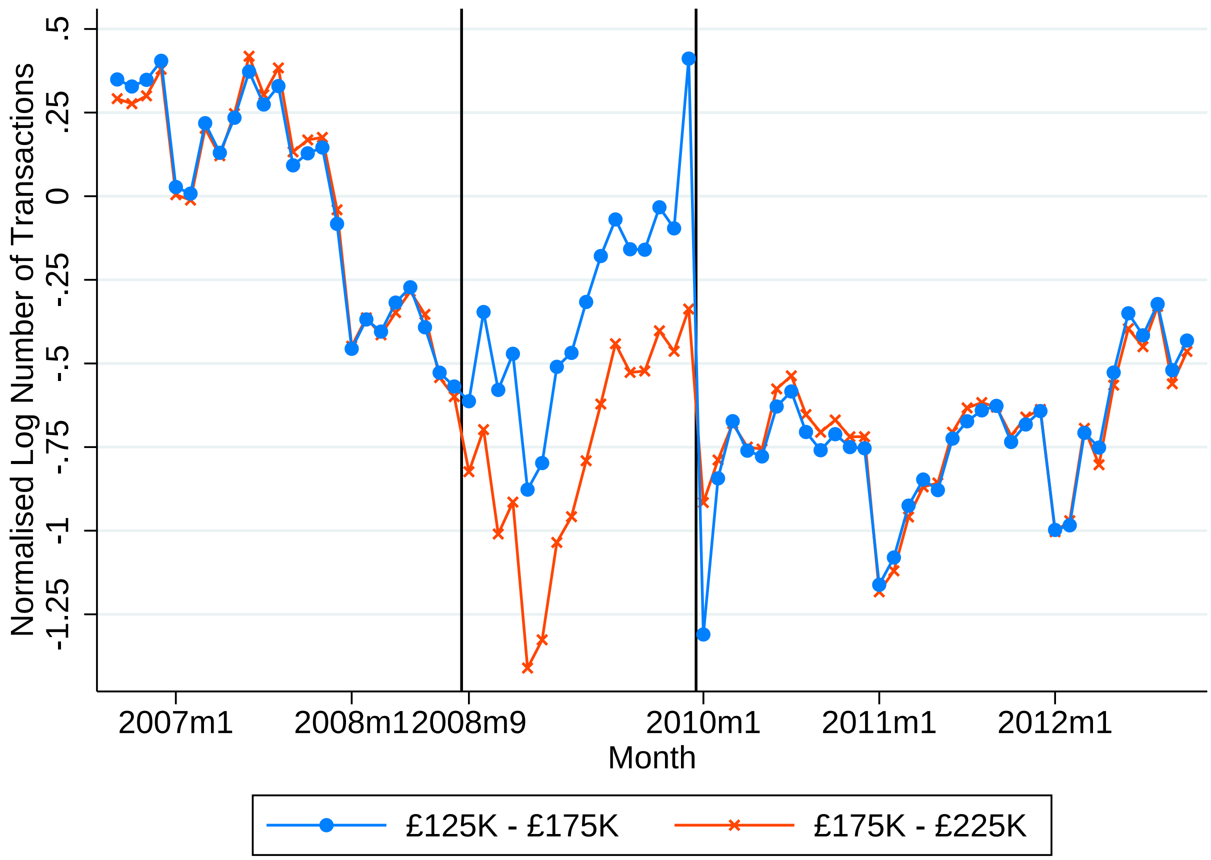


Notes: The figures show the weekly number of transactions around the end of the stamp duty holiday on 31 December 2009. Panel A shows the number of transactions taking place between 2009w27 and 2010w26 in the treated price range £125,000 – £175,000 (blue circles) alongside the number of transactions in the price ranges £75,000 – £125,000 (orange crosses) and £175,000 – £225,000 (green diamonds). Panel B shows the number of transactions taking place in the treated price range (£125,000 – £175,000) around the end of the stamp duty holiday, 2009w27 to 2010w26 (blue circles) as well as 1 year earlier (orange crosses) and 2 years earlier (green diamonds). Panel C shows the same price ranges as in panel A, but using data from 1 year earlier. Similarly, panel D shows the same price ranges as in panel A, but using data from 2 years earlier. The solid vertical line is placed at the end of the year (which at the end of 2009 is the end of the stamp duty holiday) and the dashed vertical lines demarcate the last 3 weeks of the year and the first 10 weeks of the year, which are the excluded range for the counterfactual estimates. The counterfactual is estimated according to (1.5.1):

$$c_w = \sum_{j=0}^{7} \beta_j (z_w)^j + \eta I \{w \in \text{end of month}\} + \sum_{k=\bar{w}^-}^{\bar{w}^+} \gamma_k I \{w = k\} + \mu_w$$

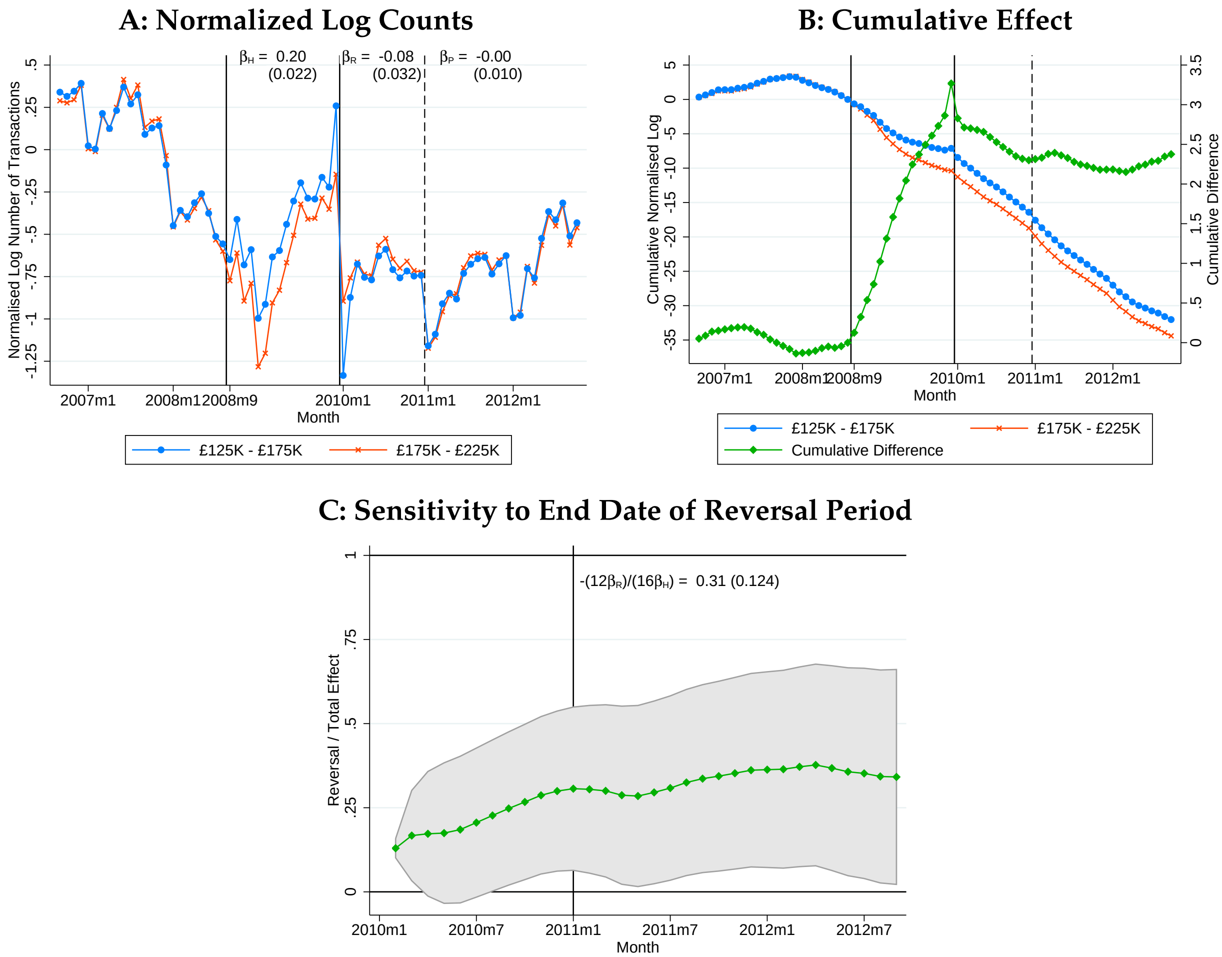
where c_w is the number of transactions in week w and z_w is the distance of week w from the end of 2009. The second term is a fixed effect for weeks at the end of the month (which feature heavier trading in every month), while the third term excludes weeks in the excluded range $(\bar{w}^-, \bar{w}^+)$. Each picture shows the difference-in-bunching estimate corresponding to the choice of treatment (blue circles) and control groups (orange crosses and green diamonds) depicted in the picture. The DiD estimate is the difference between the (normalized) bunching in the treatment group and the average bunching in the two control groups.

FIGURE 11. EFFECTS OF THE STAMP DUTY HOLIDAY STIMULUS: NAIVE DIFF IN DIFF



Notes: The figure shows how the level of housing market activity changed over time in the price range affected by the stamp duty holiday (£125,000 - £175,000) and the neighbouring price range £175,000 - £225,000. The figure shows the normalized log monthly number of transactions defined as the log of the number of transactions in that month minus the average of the log of the number of transactions in the 24 months leading up to the start date of the Stamp Duty Holiday (September 2006 - August 2008).

FIGURE 12. EFFECTS OF THE STAMP DUTY HOLIDAY STIMULUS: ADJUSTING FOR BUNCHING

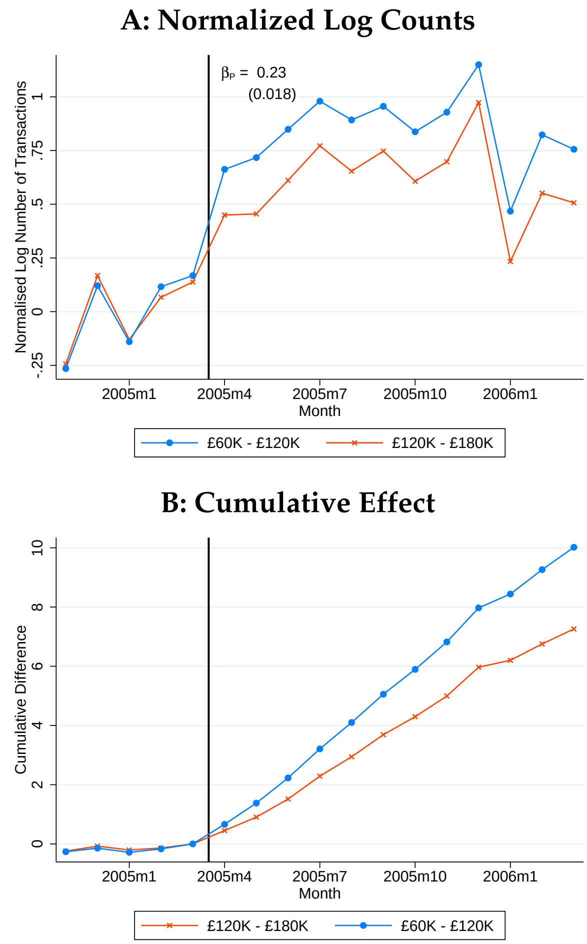


Notes: The figure shows the effect of the stamp duty holiday stimulus on housing market activity using the price range £125,000 - £175,000 as the treated price range and £175,000 - £225,000 as the control price range. However, all counts are adjusted for price manipulation using bunching estimates by moving excess transactions at £125,000 to prices between £125,000 and £150,000 and moving excess transactions at £175,000 to prices between £175,000 and £200,000. Panel A shows the normalized log monthly number of transactions defined as the log of the number of transactions in that month minus the average of the log of the number of transactions in the 24 months leading up to the start date of the Stamp Duty Holiday (September 2006 - August 2008). Superimposed on that are our estimates of β_H, β_R and β_P from the regression

$$\begin{aligned} n_{it} &= \alpha_0 Pre_t + \alpha_H Hol_t + \alpha_R Rev_t + \alpha_P Post_t + \alpha_T Treated_i \\ &\quad + \beta_H Hol_t \times Treated_i + \beta_R Rev_t \times Treated_i + \beta_P Post_t \times Treated_i + \nu_{it} \end{aligned}$$

where n_{it} is the log of the monthly number of transactions Pre_t is a dummy for the pre-period September 2006–August 2008 inclusive, Hol_t is a dummy for the stamp duty holiday period September 2008–December 2009, Rev_t is a dummy for the post-holiday reversal period January–December 2010 inclusive, and $Post_t$ is a dummy for the later months January 2011–October 2012 inclusive. $Treated_i$ is a dummy for the treated price range and finally ε_{it} is an error term. Panel B shows the cumulative sum of the normalized log counts in panel A (blue dots and orange crosses) as well as the cumulative sum of the differences between the treatment and control groups (green diamonds). Panel C shows how the proportion of the total effect of the stamp duty holiday that is undone by reversal after the end of the holiday changes as we use different months as the first month after the effect is gone. Specifically, it shows $(\Sigma_t Rev_t \times \beta_R) \,/\, (16\beta_H)$ as the end date of the period used to define Rev_t changes. The vertical line is at our preferred choice for the first month of $Post_t$, January 2011, which gives an estimate of the proportion of the total effect undone by reversal of 0.31 (0.124).

FIGURE 13. EFFECTS OF THE PERMANENT REFORM: ADJUSTING FOR BUNCHING



Notes: The figure shows the effect of the permanent tax cut of March 2005 when the bottom notch was moved from £60,000 to £120,000 on housing market activity using the price range £60,000 - £120,000 as the treated price range and £120,000 - £180,000 as the control price range. However, all counts are adjusted for price manipulation using bunching estimates by moving excess transactions at £60,000 to prices between £60,000 and £85,000 and moving excess transactions at £120,000 to prices between £120,000 and £145,000. Panel A shows the normalized log monthly number of transactions defined as the log of the number of transactions in that month minus the average of the log of the number of transactions in the 5 months we have data for leading up to the date of the reform (November 2004 - March 2005). Superimposed on that is our estimates of β_P from the regression

$$n_{it} = \alpha_0 Pre_t + \alpha_P Post_t + \alpha_T Treated_i + \beta_P Post_t \times Treated_i + \nu_{it}$$

where n_{it} is the log of the monthly number of transactions Pre_t is a dummy for the pre-period November 2004–March 2005 inclusive, $Post_t$ is a dummy for the months after the reform April 2005–March 2006 inclusive. $Treated_i$ is a dummy for the treated price range and finally ε_{it} is an error term. Panel B shows the cumulative sum of the normalized log counts in panel A (blue dots and orange crosses).

1.7. Appendix

1.7.1. A Matching Frictions Model of the Housing Market. A key feature of the competitive housing market model is that excess bunching and holes around notch points reflect real demand responses (as opposed to price incidence) and therefore reveal the elasticity of real housing demand. This section shows that the same qualitative effects on the house price distribution can be generated by bargaining between buyers and sellers in a model with matching frictions.[28] In this model, bunching responses reflect the bargaining power of buyers versus sellers.

Consider a specific match where the buyer has valuation B_v and the seller has valuation S_v of the property. Considering a flat transaction tax t (remitted by the buyer), the buyer's surplus from trading at the before-tax house price h_v is equal to $B_v - (1+t)\, h_v$ and the seller's surplus is equal to $h_v - S_v$. The necessary and sufficient condition for a trade to take place is that there exists a price such that both traders obtain a positive surplus, i.e. we must have $S_v \leq \frac{B_v}{1+t}$.

The buyer and seller engage in Nash bargaining with bargaining power β for the buyer and $1-\beta$ for the seller. The agreed before-tax price h_v^* maximizes $W = \left[B_v - (1+t)\, h_v\right]^{\beta} \left[h_v - S_v\right]^{1-\beta}$, which yields

$$h_v^* = \beta S_v + (1-\beta)\, \frac{B_v}{1+t}. \tag{1.7.1}$$

Hence, conditional on trading, the transaction tax reduces the house price h_v^*, with the strength of the price effect being proportional to the bargaining power of the seller $1-\beta$. This means that we can characterize the effects of the transaction tax t in the following way. House transactions that were desirable to the buyer and seller in the absence of transaction taxes but sufficiently close to the indifference margin for both ($B_v/\,(1+t) < S_v \leq B_v$) will no longer occur (extensive response). House transactions that continue to be desirable in the presence of transaction taxes ($S_v \leq B_v/\,(1+t)$) will occur at lower prices according to equation (1.7.1). Assuming a smooth distribution of matches S_v, B_v and bargaining power β, captured by a density distribution $f\,(S_v, B_v, \beta)$, there will be a smooth distribution of traded house prices under the flat transaction tax t.

[28]Our matching frictions model for the housing market is conceptually similar to the labor market model used by Kleven *et al.* (2013) to study income taxes and migration.

Consider now the introdution of a notch Δt in the transaction tax at the cutoff house price $\overline{h}_v$. Under the notched tax schedule and Nash bargaining between the buyer and seller, the agreed house price h_v is picked to maximize

$$W = \left[B_v - \left(1 + t + \Delta t \cdot I\left\{h_v > \overline{h}_v\right\}\right) h_v\right]^\beta \left[h_v - S_v\right]^{1-\beta}. \tag{1.7.2}$$

In general, solving this bargaining problem requires us to solve for the best price point within each tax bracket (below and above $\overline{h}_v$) and then pick the candidate solution that yields the largest welfare W. Trades that would occur below $\overline{h}_v$ under the baseline flat tax are clearly unaffected by the notch and continue to feature house prices given by (1.7.1). On the other hand, trades that would occur above $\overline{h}_v$ under the baseline flat tax are affected by the notch. To see how these trades are affected, note first that any trade occurring stricly above the cutoff must satisfy the interior pricing condition (1.7.1) with the $1 + t$ replaced by $1 + t + \Delta t$. This allows us to distinguish between three cases.

First, some transactions just above $\overline{h}_v$ under the baseline tax rate t would have an interior solution below $\overline{h}_v$ under the larger tax rate $t + \Delta t$ (based on eq. (1.7.1) at tax rate $1 + t + \Delta t$). This is inconsistent with an interior solution in either bracket, and so these transactions bunch at the cutoff. Second, some transactions that were taking place in a region $(\bar{h}_v, \bar{h}_v + \Delta\bar{h}_v)$ in the absence of the notch and that would be just above $\overline{h}_v$ under an interior solution at the new tax rate $t + \Delta t$ (again based on eq. (1.7.1) at tax rate $1 + t + \Delta t$) also bunch at the cutoff. For such transactions, a small move to the cutoff provides a discrete gain to the buyer and only a marginal loss to the seller, yielding a larger value of W than at the interior location. Of course, for such a move to be possible, it must be the case that the seller still receives positive surplus, so only those transactions for which $S_v \leq \bar{h}_v$ will bunch. Given a smooth distribution of matches (S_v, B_v), there will be marginal bunching transactions such that welfare at the cutoff $\overline{h}_v$ is precisely equal to welfare at the best interior location above the notch $\bar{h}_v^I$. In the interval $\left(\overline{h}_v, \bar{h}_v^I\right)$ all transactions with $S_v \leq \bar{h}_v$ move to the threshold and so we get a hole in the price distribution there. The width of this hole depends on bargaining power and converges to zero as the bargaining power of the buyer β converges to zero.[29] Third and finally, transactions above $\bar{h}_v^I$ under

[29]These marginal transactions satisfy

$$\left(B_v - (1+t)\,\bar{h}_v\right)^\beta \left(\bar{h}_v - S_v\right)^{1-\beta} = \left(B_v - (1+t+\Delta t)\,\bar{h}_v^I\right)^\beta \left(\bar{h}_v^I - S_v\right)^{1-\beta} \tag{1.7.3}$$

an interior solution at the new tax rate $t + \Delta t$ are associated with a larger W at the new interior solution than at the cutoff. For those transactions, we get a downward price shift within the upper bracket.

This characterization applies only to matches for which a trade is still beneficial. The notch will also create extensive responses above the cutoff as house transactions that were desirable to the buyer and seller under the flat tax but close enough to the indifference margin for both $(B_v / (1+t+\Delta t) < S_v \leq B_v / (1+t))$ and which cannot take place with positive surplus at the notch (as $\bar{h}_v < S_v$) will no longer occur. Nevertheless, as in the competitive model, extensive responses are negligible *just* above the cutoff. Trades that would occur at a price $h_v \in (\bar{h}_v, \bar{h}_v + \Delta\bar{h}_v)$ in the absence of the notch (but have a negative surplus under the higher tax, i.e. $B_v / (1+t+\Delta t) < S_v \leq B_v / (1+t))$ must have a positive surplus under the lower tax such that $S_v \leq h_v \leq B_v / (1+t)$. In the presence of the notch, for those trades to take place at the cutoff price $\bar{h}_v$ it must be the case that $S_v \leq \bar{h}_v \leq B_v / (1+t)$. Together these conditions imply that those trades cannot achieve positive surplus by bunching at the notch whenever $S_v \in NT = (\bar{h}_v, h_v)$. As the price absent the notch h_v converges to $\bar{h}_v$ from above, we see that the no-trade set NT becomes empty and so there is no extensive margin response just above the threshold. Finally, note that the presence of the notch could shift the distribution of buyer and seller matches S_v, B_v above the notch, for example, by inducing buyers and sellers with valuations that put them near the notch to continue searching in order to find another match. We supress these effects for simplicity, but again, they will be negligible *just* above the notch.

The characterization above is analogous to the characterization for the competitive model, with the bargaining power parameter β in the bargaining model playing the role of the demand elasticity α in the competitive model. A graphical illustration similar to Figure 1 is also possible. Figure 14 shows the direct analog of panel A of Figure 1 for the case of the bargaining model, and shares all of its qualitative features. The density diagrams in panels C-D of Figure 1 can also be reinterpreted in terms of the bargaining model, with panel C depicting the intensive margin effects on the house price distribution for the full distribution of βs and panel D incorporating the extensive margin effects. We can summarize the bargaining model's predictions as follows

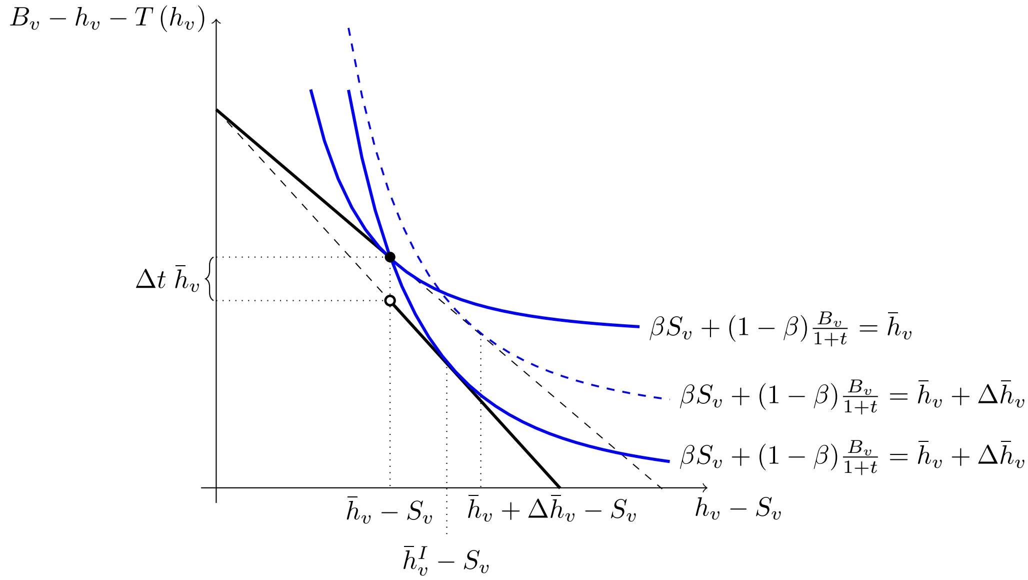
where $\bar{h}_v^I = \beta S_v + (1-\beta)\frac{B_v}{1+t+\Delta t}$ and $\bar{h}_v + \Delta\bar{h}_v = \beta S_v + (1-\beta)\frac{B_v}{1+t}$. From this we can also immediately see that the width of the hole converges to 0 as the bargaining power parameter β converges to 0.

PROPOSITION 3 (**Notches with Matching Frictions**). *A transaction tax featuring a notch at a property value $\bar{h}_v$ at which the proportional tax rate jumps from t to $t + \Delta t$ induces*

***(i)** an **intensive margin response** as matches in the house price range $(\bar{h}_v, \bar{h}_v + \Delta\bar{h}_v)$ for which $S_v \leq \bar{h}_v$ bunch at the threshold $\bar{h}_v$, where the width of the bunching segment $\Delta\bar{h}_v$ is monotonically increasing in the bargaining power parameter β as characterized by equation (1.7.3); and*

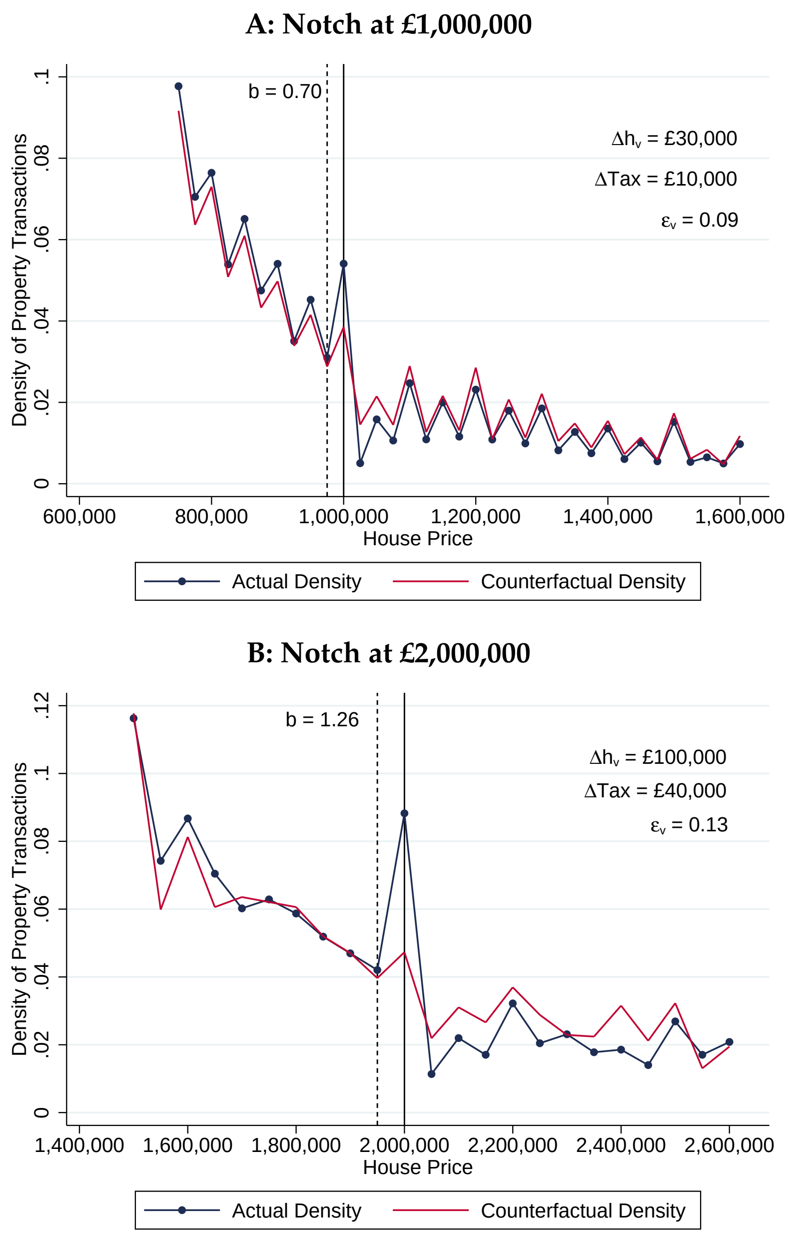
***(ii)** an **extensive margin responses** as matches in the house price range $h_v \in (\bar{h}_v, \infty)$ for which $B_v / (1 + t + \Delta t) < S_v \leq B_v / (1 + t)$ and $S_v \in NT = (\bar{h}_v, h_v)$ choose not to trade. The extensive response converges to zero just above the cutoff as the set NT converges to the empty set as $h_v \to \bar{h}_v^+$.*

FIGURE 14. BUDGET SET DIAGRAM FOR BARGAINING MODEL



Notes: The budget set diagram depicts the Nash product as in equation (1.7.2) and the budget set of feasible allocations under the notched tax schedule in the space of net of tax surpluses (i.e. $(B_v - h_v - T(h_v), h_v - S)$-space) and illustrates intensive responses among individuals with heterogeneous valuations $\{B_v, S_v\}$, but a specific bargaining power β. The notch creates bunching at the cutoff $\overline{h}_v$ by all individuals in a preference range $\beta S_v + (1-\beta)\frac{B_v}{1+t} \in [\bar{h}_v, \overline{h}_v + \Delta\overline{h}_v]$, who would have bargained prices on the segment $[\overline{h}_v, \overline{h}_v + \Delta\overline{h}_v]$ in the absence of the notch. The marginal bunching match is indifferent between the notch point $\overline{h}_v$ and the best interior location $\bar{h}_v^I$. No individual is willing to locate between $\overline{h}_v$ and $\bar{h}_v^I$, and hence this range is completely empty. This figure is the direct analog of panel A of figure 1, and shares all its qualitative features.

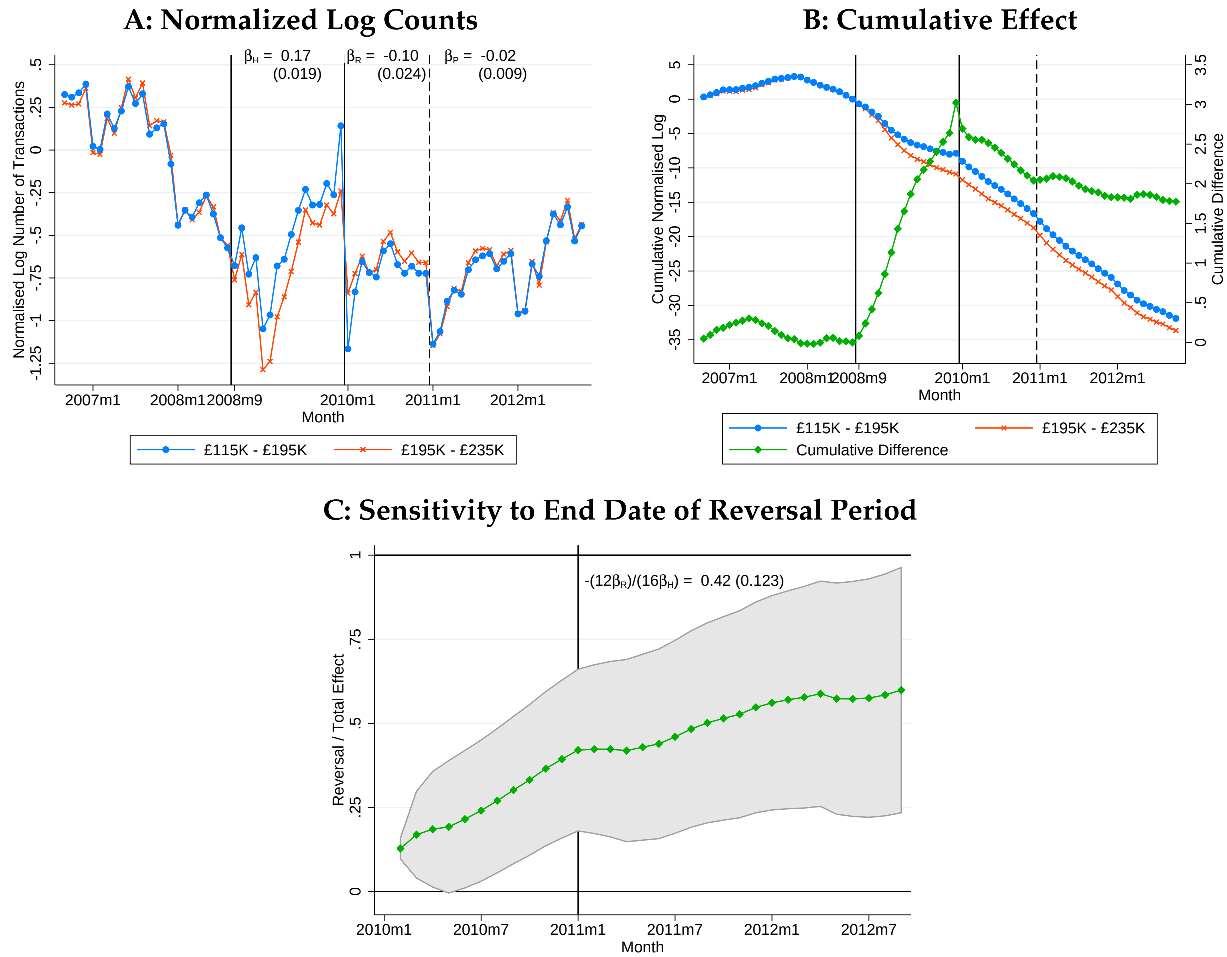
FIGURE 15. BUNCHING AND HOLES AROUND THE HIGHEST NOTCHES



Notes: The figure shows the observed density of property transactions (blue dots) and the density of property transactions in the year leading up to the introduction of the notch (red line) around the notches for very high value properties. The vertical dashed lines denote the upper and lower bounds of the excluded region around the notch. The upper bound of the excluded region is chosen as the point where the observed density changes slope from positive to negative. Panel A shows the notch at £1,000,000 introduced on 6 April 2011 where the tax liability jumps by £10,000 (from 4% to 5% of the transaction price) with both densities aggregated up to bins £25,000 wide. Panel B shows the notch at £2,000,000 introduced on 22 March 2012 where the tax liability jumps by £40,000 (from 5% to 7% of the transaction price) with both densities are aggregated up to bins £50,000 wide. b is our estimate of the excess mass just below the notch scaled by the average counterfactual frequency in the excluded range and m is our estimate of the missing mass above the notch scaled by the average counterfactual frequency in the excluded range. $m - b$ is our estimate of the difference between the missing mass and the bunching mass. The figures also show the average house value change created by the notch, and the tax liability change at the notch.

1.7.2. Additional Figures & Tables.

FIGURE 16. EFFECTS OF THE STAMP DUTY HOLIDAY STIMULUS: DIFF IN DIFF WITH WIDER TREATMENT RANGE



Notes: The figure shows the effect of the stamp duty holiday stimulus on housing market activity using the price range £115,000 - £195,000 as the treated price range and £195,000 - £235,000 as the control price range. Panel A shows the normalized log monthly number of transactions defined as the log of the number of transactions in that month minus the average of the log of the number of transactions in the 24 months leading up to the start date of the Stamp Duty Holiday (September 2006 - August 2008). Superimposed on that are our estimates of β_H, β_R and β_P from the regression

$$\begin{aligned} n_{it} &= \alpha_0 Pre_t + \alpha_H Hol_t + \alpha_R Rev_t + \alpha_P Post_t + \alpha_T Treated_i \\ &\quad + \beta_H Hol_t \times Treated_i + \beta_R Rev_t \times Treated_i + \beta_P Post_t \times Treated_i + \nu_{it} \end{aligned}$$

where n_{it} is the log of the monthly number of transactions Pre_t is a dummy for the pre-period September 2006–August 2008 inclusive, Hol_t is a dummy for the stamp duty holiday period September 2008–December 2009, Rev_t is a dummy for the post-holiday reversal period January–December 2010 inclusive, and $Post_t$ is a dummy for the later months January 2011–October 2012 inclusive. $Treated_i$ is a dummy for the treated price range and finally ε_{it} is an error term. Panel B shows the cumulative sum of the normalized log counts in panel A (blue dots and orange crosses) as well as the cumulative sum of the differences between the treatment and control groups (green diamonds). Panel C shows how the proportion of the total effect of the stamp duty holiday that is undone by reversal after the end of the holiday changes as we use different months as the first month after the effect is gone. Specifically, it shows $(\beta_R \Sigma_t Rev_t)$ / $(16\beta_H)$ as the end date of the period used to define Rev_t changes. The vertical line is at our preferred choice for the first month of $Post_t$, January 2011, which gives an estimate of the proportion of the total effect undone by reversal of 0.42 (0.123).

TABLE 4. IMMEDIATE IMPACT OF FISCAL STIMULUS ON GDP

	Time Since Last Move		
	$<$ 1 Year	$\geq$ 1 Year	$\geq$ 5 Years
Panel A: Moving-Related Household Spending			
Repairs & Improvements	3,153	1,707	1,447
Furnishings	2,912	817	751
Appliances	153	87	100
Other Durables	426	434	436
Total Expenditure	6,644	3,043	2,734
Difference Movers - Stayers		3,600	3,909
Difference Movers - Stayers (% of house value)		1.57	1.70
Estate Agent Commissions (% of house value)		1.98	1.98
Other Commissions (% of house value)		1.24	1.24
Impact of Purchase on Expenditure: ϕ (% of house value)		4.79	4.92
Panel B: Immediate Impact of Policy on GDP			
Impact of Policy on GDP per £ of Tax Cut ($\beta_H \times \phi$) using $\beta_H = 0.20$		0.96	0.98
Impact of Policy on GDP per £ of Tax Cut ($\beta_H \times \phi$) using $\beta_H = 0.17$		0.81	0.84

Notes: The table shows estimates of the immediate impact of the stamp duty holiday stimulus on GDP. Using the UK Living Costs and Food Survey from 2011, the first 7 rows of panel A present estimates of moving-related spending on repairs, renovations, furnishings, appliances and other durables. To obtain only the moving-related part of these spending categories, we compare homeowners who moved within the last year (movers) to homeowners who moved more than 1 year ago or more than 5 years ago (non-movers). Row 6 shows our estimates of total moving-related spending on these categories in absolute numbers (£3,600-£3,909 depending on comparison group), while row 7 scales the estimates by the average house price of houses transacted in 2011, £230,000. Rows 8 and 9 show spending on commisions to agents, lawyers, etc. A 2011 survey by *Which? Magazine* estimates that estate agents' fees average 1.8% of the house price before VAT, or 1.98% with VAT (see http://www.which.co.uk/news/2011/03/estate-agents-fees-exposed-248666/). ReallyMoving (2012) estimates that other commissions and fees total £1,880 on average, and do not vary much with house value, so we scale this by the average value of houses bought in the range affected by the policy (£152,000). Combining rows 1-9, we reach our rough estimate of the effect of a house purchase on household spending (in % of the house price), which we denote by ϕ. This number is just below 5% independent of comparison group. In panel B we calculate the immediate impact of the policy on GDP (per £ of tax cut) as the moving-related spending triggered by the additional house transactions due to the policy. The total GDP effect is $\Delta GDP = \phi h_v^m \Delta n$ where h_v^m is the mean price of houses in the price range affected by the policy, and Δn is the number of additional house purchases induced by the policy. The foregone tax revenue is $\Delta Tax = \tau_0 h_v^m n_0$ where $\tau_0 = 1\%$ is the pre-stimulus tax rate, and n_0 is the counterfactual number of transactions. Combining these expressions, the effect of the policy is $\Delta GDP/\Delta Tax = \phi \Delta n/\left(\tau_0 n_0\right)$, where $\Delta n/n_0$ is our difference-in-differences estimate β_H in equation (1.5.2). The first row of panel B uses $\beta_H = 0.20$ as estimated in Figure 12, while the second row uses $\beta_H = 0.17$ as estimated in Figure 16.

CHAPTER 2

The Role of Firms in Workers' Earnings Responses to Taxes: Evidence From Pakistan[1]

ABSTRACT. This paper exploits employee-employer matched administrative tax data on firms and salaried workers in Pakistan to explore the underappreciated role of firms in determining how workers' taxable earnings respond to taxation. I present evidence on three ways in which firms affect workers' earnings responses. First, third-party reporting of salaries by employers makes underreporting taxable income more costly for workers and reduces evasion of the income tax. Second, firms' equilibrium salary-hours offers respond endogenously to the presence of adjustment costs in the labour market by tailoring offers to aggregate worker preferences. Third, workers learn about the tax schedule from firms' salary offers, making them more responsive to taxation both contemporaneously (by 130%) and in subsequent years (by 100%). However, while third-party reporting makes misreporting more costly, it does not eliminate it in a low tax-capacity setting: 19% of workers still underreport their salaries, leading to a loss of about 5% of tax revenue, and indicating high returns to investments in improving enforcement capacity. The large role played by firms in determining workers' earnings implies that firms need to play a central role in our analysis of income taxation in lower income countries.

[1]I would like to thank my supervisors, Henrik Kleven and Oriana Bandiera for their generous support during this project. I am grateful to Camille Landais, Gerard Padró i Miquel and Johannes Spinnewijn for numerous helpful discussions, to Miguel Almunia, Giuseppe Berlingieri, Steve Bond, Michael Devereux, Ulrich Doraszelski, Greg Fischer, Anders Jensen, Torsten Persson and Mazhar Waseem and numerous seminar participants for useful suggestions, and to Ali Arshad Hakeem, Samad Khurram, Jawad Abbasi and Ijaz Nabi for their help with the data and their enthusiasm. Financial support from the International Growth Centre, Pakistan Programme made this project possible. The plethora of remaining errors is mine alone.

2.1. Introduction

The development of the capacity of the state to raise taxes is central to the process of economic development, but the public finance literature has been largely silent on the issue, either tending to assume that taxes can be perfectly and costlessly enforced or taking evasion and administrative costs as given (Besley & Persson, 2011, 2013). A recent literature has suggested that third-party reporting of tax bases to the tax authority, particularly by firms, is key to understanding the government's capacity to enforce taxes (Kleven *et al.*, 2009; Pomeranz, 2013). In the case of the personal income tax, third-party reporting of salaries and withholding of income tax by employers form the bedrock of the enforcement regime in modern tax administrations. Historically, the first successful modern income taxes in both the UK and the US featured withholding of income tax on the salaries of civil servants (Slemrod, 2008). Today, all 34 OECD countries require employers to report their employees' salaries and all except France require employers to withhold income tax on their employees, collecting over 75% of personal income tax revenues through withholding (OECD, 2013).

More generally, the role of firms in the study of taxation has been underplayed. Kopczuk & Slemrod (2006) appeal for firms to be central to models of taxation with imperfect enforcement, noting firms' key roll in collecting and enforcing taxes.[2] Firms may also play a broader role in determining the way that workers' reported taxable incomes respond to taxes. In the presence of adjustment costs in the labour market (such as costly search), workers are not simply paid their marginal product (obviating a role for firms). Instead, salary earnings are the outcome of a matching process to which firms are central (see, for example, Rogerson *et al.*, 2005 and Manning, 2011 for surveys, and Chetty *et al.*, 2011 for an application to taxation). Similarly, if workers face information frictions preventing them from perceiving the tax schedule they face accurately, firms' behaviour during the salary determination process can convey useful information on the tax schedule and this can influence how workers respond to taxation, both in their salary and non-salary earnings.

This paper provides evidence on both issues in the context of the taxation of salaried workers in Pakistan. I am able to leverage four key advantages of my data and setting

[2]For example, in the United States, 82% of federal tax revenues are remitted by firms (Slemrod, 2008)

in order to bring evasion under third-party reporting and withholding, and the role of firms in determining workers' earnings responses to taxation into sharp relief. First, I have benefited from being able to exploit large administrative datasets of tax records from the Federal Board of Revenue (FBR) on the universe of tax-registered firms and workers in Pakistan.[3] The large sample sizes on the universe of taxpayers (roughly a million records per year) and the existence of administrative identifiers permitting the linkage of individuals and firms across datasets and years allow me to overcome the problems of attrition, non-response, and measurement error that typically plague studies using survey data (Card *et al.*, 2010), especially in developing countries (De Mel *et al.*, 2009). Second, to my knowledge, this paper is the first to be able to directly study third party reporting of wages by firms and workers in an employer-employee matched dataset from a developing country.[4] Kumler *et al.* (2013) and Niehaus & Sukhtankar (2013) study similar issues of misreporting of wages, but are unable to match individual-level observations of salaries and so study discrepancies between the distributions of wages reported by firms and by workers.[5]

Third, the tax schedule for the salaried employees I study features a large number of tax brackets (between 17 and 20 during the period I study) and the kinks in firms' and workers' choice sets induced by the jumps in the marginal tax rate at the bracket thresholds provide multiple compelling sources of quasi-experimental variation in tax incentives. These discontinuities in marginal tax rates generate incentives for incomes to cluster around the bracket thresholds, allowing me to non-parametrically identify behavioural responses to marginal tax rates using a bunching approach (Saez, 2010; Chetty *et al.*, 2011; Kleven & Waseem, 2013). Fourth, in Pakistan, as in other low income countries, it is common for individuals to diversify their income sources. This means that my data contains a considerable number of observations on individuals with significant amounts of non-salary income. In combination with the large number of tax brackets, this provides a

[3]By its very nature, informal economic activity is not captured in government records and so does not form a part of this study.

[4]Gerard & Gonzaga (2013) study labour informality and unemployment insurance in Brazil using an employee-employer matched dataset, but do not have independent reports of wages from workers and from firms. Carrillo *et al.* (2013) exploit experimental variation in cross-checks of third party reports of business to business transactions in Ecuador.

[5]Similarly, Fisman & Wei (2004) study discrepancies in the distribution across product categories of reported imports and exports to detect tax evasion.

unique opportunity as worker-firm matches are likely to be responding to multiple kinks simultaneously, allowing me to disentangle responses in salary and non-salary incomes, and responses by firms and by workers.

To guide the empirical analysis, I set out a model in which firms and workers interact to determine salaries, and workers independently set their non-salary earnings. The model has three key features. First, workers can underreport their earnings at a cost. Second, the salaried labour market features adjustment costs. Third, some workers face information frictions preventing them from responding optimally to the incentives generated by the tax schedule. In turn, this provides three channels through which firms can affect workers' taxable earnings. Third-party reporting of salaries raises the cost of misreporting salaries; firms' salary offers can respond to the presence of adjustment costs; and firms' salary offers can convey information about the tax schedule to prospective workers. The model generates predictions regarding the extent of bunching of salary and taxable incomes around kinks in the tax schedule, which I then take to the data.

I present 5 sets of empirical findings. First, I document the presence of sharp bunching of overall taxable incomes around kinks in the tax schedule, providing direct evidence of behavioural responses to taxation in a lower income country context. Conceptually, these taxable income responses can be comprised of evasion responses, real earnings responses, and earnings shifting responses, so the next findings provide evidence on each of those.

Second, unilateral salary underreporting by workers is widespread. 19% of workers underreport their salary, underreporting it by an average of 16%. This leads to 4% of salary income going untaxed, or at least 5% of the tax revenue from salaried employees being lost. Consistent with the model, the prevalence and level of misreporting is positively correlated with the marginal tax rate faced by the worker, and with the share of the worker's total income that is self-reported. This misreporting is orders of magnitude larger than the available evidence from high income countries indicates. For instance, Kleven *et al.* (2011) find misreporting by only 1.3% of workers, amounting to 0.2% of income in Denmark. Moreover, since I am unable to detect misreporting that workers and firms collude in, the results presented here are only a loose lower bound on misreporting.

Third, I document *firm bunching* of salary incomes around kinks in the tax schedule amongst workers who do not face a kink in their budget set at these statutory kinks.

Consistent with the predictions of the model, this provides direct evidence of firm-level responses to aggregate worker preferences, and of the presence of significant adjustment costs on the part of workers. In contrast to the finding by Chetty *et al.* (2011) of "aggregate bunching" around a kink in Denmark, a highly unionized labour market, in Pakistan the role of unions is negligible, and so I provide the first direct evidence of firms (rather than unions) aggregating workers' preferences.

Fourth, I document the presence of significant *double bunching*—individuals with salaries at one kink, and taxable incomes at a different kink. The model predicts that taxable income bunching is reduced when individuals are constrained by adjustment costs to accept a suboptimal salary income, and hours spent on salaried and non-salaried work are imperfectly substitutable. As a result of firm bunching, individuals with salaries at a kink are disproportionately likely to be facing significant adjustment costs, so the presence of significant double bunching indicates that workers are able to shift earnings between salary and non-salary income relatively easily. This is evidence of shifting between salary and non-salary income within the personal income tax base, rather than across bases as has more traditionally been studied (Gordon & Slemrod, 2000).

Fifth, I pursue both cross-sectional and event study methodologies to provide evidence that workers face information frictions preventing them from responding to the kinks in the income tax schedule, and that firm offers convey information about the tax schedule, increasing workers' overall responsiveness. I find that firm bunching increases worker responsiveness by around 130% in the year a worker receives a salary at a kink, and by 100% in subsequent years. These large results are similar in magnitude to those in Chetty *et al.* (2013) who find that moving to a high information neighbourhood from a median information neighbourhood roughly doubles workers' propensity to bunch at the refund-maximising kink in the Earned Income Tax Credit (EITC) in the United States.

This paper contributes to 3 literatures. First, there is a large literature on the determinants of tax evasion (see Andreoni *et al.*, 1998 and Slemrod & Yitzhaki, 2002 for surveys) and on estimating the extent of tax evasion (see Slemrod, 2007 and Slemrod & Weber, 2012 for surveys). This literature has been plagued with methodological and measurement issues, and this paper contributes to an emerging literature using discrepancies between two reports on the same tax base to study evasion (Fisman & Wei, 2004; Kumler *et al.*, 2013;

Niehaus & Sukhtankar, 2013; Zucman, 2013). There is also a small literature studying the effects of third-party reporting and withholding on tax evasion, either in rich-country contexts, or studying evasion of taxes on firms, rather than workers (Yaniv, 1988; Slemrod *et al.*, 2001; Kleven *et al.*, 2011; Carrillo *et al.*, 2012; Pomeranz, 2013).

Second, a recent public finance literature posits that optimization frictions can account for the large discrepancies between microeconometrically estimated labour supply (or more generally, taxable income) elasticities and macro estimates (Chetty *et al.*, 2011; Chetty, 2012). Jones (2012) and Gelber *et al.* (2013) also study the implications of optimization frictions for the dynamics of adjustment to policy, finding a large role for adjustment costs. Of particular note here, Kleven & Waseem (2013) find that elasticities unattenuated by optimisation frictions are between 5 and 10 times larger than those implied by observed bunching behaviour at notches in the tax schedule in Pakistan. A second literature focuses specifically on information frictions, finding substantial effects of tax salience on demand elasticities (Chetty *et al.*, 2009; Finkelstein, 2009) and even on political instability (Cabral & Hoxby, 2012). A number of papers in this literature also consider the effectiveness of programmes that aim to increase responsiveness through the provision of information (Duflo & Saez, 2003; Liebman & Luttmer, 2011; Chetty & Saez, 2013) with mixed findings. Liebman & Zeckhauser (2004); Bernheim & Rangel (2009); Chetty *et al.* (2009); Mullainathan *et al.* (2012); Spinnewijn (2013b,a) also study the theoretical implications of misperception of choice sets for welfare and optimal policy.

Third, this paper contributes to a burgeoning literature on public finance and development (see Besley & Persson, 2013 for a survey) and particularly to work using administrative microdata and quasi-experimental methods to evaluate tax policy in developing countries (Kleven & Waseem, 2013; Best *et al.*, 2013; Kumler *et al.*, 2013; Pomeranz, 2013).

This paper proceeds as follows. Section 2.2 presents a model of salary determination by firms and workers and of non-salary earnings choices by workers to guide the empirial analysis. Section 2.3 presents the Pakistani context and the data I use. Section 2.4 presents the results on overall taxable income bunching, and evidence on its three constituent parts: evasion (2.4.2), real responses (2.4.3) and income shifting (2.4.4). Section 2.5 presents evidence that workers learn about the tax schedule from their interactions with employers. Finally, section 2.6 concludes.

2.2. Conceptual Framework

This section develops a simple, stylized model of the determination of salaries by firms and workers and of workers' joint choice of salary and non-salary earnings. The model has three key features. First, workers can underreport their earnings at a cost. Second, the salaried labour market features adjustment costs. Third, some workers face information frictions preventing them from responding optimally to the incentives generated by the tax schedule. While the model is extremely stylized, it captures the relevant features of the environment and serves to guide the empirical analysis by generating predictions regarding the extent of bunching of salary and taxable incomes around kinks in the tax schedule which I then take to the data.

Workers. Individuals have quasilinear preferences over consumption c, a CES aggregate of hours spent on salaried work l and hours spent on non-salaried work q, and reported salary and non-salary earnings, $\hat{s}$ and $\hat{n}$:

(2.2.1)
$$U\left(c,l,q,\hat{s},\hat{n}\right)=c-\frac{\left(\alpha+\beta\right)^{1-\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)}}{1+\frac{1}{\varepsilon}}\left[\alpha\left(\frac{l}{\alpha}\right)^{\sigma}+\beta\left(\frac{q}{\beta}\right)^{\sigma}\right]^{\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)}-e_0 I\left\{\hat{s}<s\right\}-e\left(s-\hat{s},n-\hat{n}\right)$$

Individuals have heterogeneous tastes for working parameterized by $\alpha>0$ and $\beta\geq 0$ capturing heterogeneity in abilities and disutilities of labour supply. A fraction η of individuals have $\beta=0$ implying that these individuals earn only salary income, while the remaining fraction with $\beta>0$ have access to a linear production technology allowing them to produce the same output as firms.[6] However, in contrast to firms, they are able to costlessly adjust their labour input. Denote the cdf of α,β as $G\left(\alpha,\beta\right)$ with corresponding density $g\left(\alpha,\beta\right)$.

Misreporting. Workers declare their salary income $\hat{s}$ and their non-salary income $\hat{n}$ to the tax authority, and pay taxes on their taxable income $\hat{z}=\hat{s}+\hat{n}$ according to a piecewise linear tax schedule $T\left(\hat{z}\right)$ featuring two kinks at threshold taxable income levels K_1,K_2 at which the marginal tax rate jumps up from τ_{j-1} to $\tau_j>\tau_{j-1}$, $j=1,2$. The worker's true

[6]The assumption that η is independent of α simplifies the exposition, but can easily be relaxed.

earnings are $z = s + n$, the sum of their salary earnings $s = wl$ at wage rate w and their non-salary earnings $n = pq$, where p is the price of final output.[7]

However, at a cost, the workers can misreport their incomes, reporting $\hat{s} < s$ and/or $\hat{n} < n$. The cost of misreporting has two parts. First, due to third-party reporting of salaries by employers, the tax authority sometimes cross-checks the employee's and the employer's reports, and so reporting a salary of $\hat{s} < s$ carries a fixed cost of $e_0 > 0$. Since the probability that the two salary reports are cross-checked, and the reliability of the employer's report vary, workers are heterogeneous in the fixed cost they face, with the fixed cost distributed according to $D(e_0)$. Second, the cost of misreporting depends on the misreported amounts, where I assume that $e(\cdot,\cdot)$ is increasing in both arguments and convex, with $\partial^2 e/\partial(n-\hat{n})\,\partial(s-\hat{s}) > 0$, and that $e(0,0) = 0$. I also assume that workers who are more porductive in self-employment (higher β) are also better able to convincingly misreport their income, so that $\partial e/\partial\beta < 0$, but that this effect has diminishing returns, so that $\partial^2 e/\partial(n-\hat{n})\,\partial\beta > 0$.

Information Frictions. I remain agnostic about the precise mechanism through which some individuals have failed to learn about the full tax schedule, and model information frictions in a reduced form way (Mullainathan *et al.*, 2012). I simply assume that some individuals (denoted by $\lambda = 1$) are aware of the kinks in the tax schedule and respond optimally to the full tax schedule, while the remaining individuals are naïve and behave as if the tax schedule did not feature kinks.[8] Workers' optimal salary and non-salary labour supplies are given by $s^*, n^* = \arg\max_{s,n} U$ where sophisticated workers optimize using the appropriate budget constraint $c = z - T(\hat{z})$ while naïve taxpayers fail to, instead using some other budget constraint that does not feature the kinks induced by $T(\hat{z})$.

Firms. As in Chetty *et al.* (2011), firms are modelled extremely simply as producing output according to a linear, one-factor production function employing only labour. Firms post offers consisting of a package of a number of hours worked l and a wage rate $w(l)$, and commit to these offers *before* matching with workers, giving rise to a distribution of of

[7] For simplicity, I assume individuals have no non-labour income. In the case of quasi-linear utility, this makes no diference to the results. With more general utility, this will introduce income effects, but the qualitative conclusions remain unchanged.

[8] This could be for a variety of reasons. For example, it may be too costly for individuals to process the necessary information (Sims, 2003; Schwartzstein, 2012), they may confuse average and marginal rates (Liebman & Zeckhauser, 2004; Chetty *et al.*, 2009), or individuals may forget where the kinks are from year to year (Mullainathan, 2002).

hours offers $F^o(l)$. Furthermore, firms are unable to condition offers on the non-salary income of workers, i.e. $F^o(l|n) = F^o(l)\ \forall n$. A firm that posts a job requiring s hours thus earns profits of

$$\Pi = pl - w(l)\,l$$

Once a firm hires a worker, the worker's salary $w(l)\,l$ is also reported (truthfully) to the tax authority.

Fixed Costs of Adjustment. Workers must engage in a costly search process to match with jobs. As with the information frictions, I will remain agnostic about the precise source of these adjustment costs. Following Chetty *et al.* (2011) I assume that workers randomly sample a job offer from $F^o(l)$, and choose either to accept this job, or to pay a fixed cost ϕ, in which case I assume they find a job paying their preferred salary $s^* = w(l^*)\,l^*$ with certainty.

Equilibrium. The search process will map the distribution of posted offers $F^o(l)$ and the wage schedule $w(l)$ to a distribution of accepted salaries $\mathcal{D}[F^o, w]$ which combines the distribution of offers and the distribution of preferred hours. In order for the labour market to clear, it must be the case that the distribution of posted offers equals the distribution of accepted offers, or that $F^o(l) = \mathcal{D}[F^o(l), w(l)]$. That is, labour market equilibrium is a fixed point of $\mathcal{D}(\cdot)$. Furthermore, assuming free entry into a competitive market for final output, profits are bid down to zero and $w(l) = p\,\forall l$.

I proceed to analyze this model through a series of special cases. With the exception of the case focussing on salary misreporting, in each case, the model's equilibrium is summarized in terms of what it predicts for the degree of bunching of salary income and total taxable income at the kinks in the tax schedule, and how taxable income bunching varies with salary income. In particular, each equilibrium gives rise to a distribution of salary income $H(s)$ and a distribution of taxable income $J(z)$. Bunching in the salary income distribution is then the excess mass at the kinks

$$B_s(K_j) = H(K_j) - \lim_{s \uparrow K_j} H(s)\ \ j = 1, 2$$

and similarly bunching in the taxable income distribution is

$$B_z(K_j) = J(K_j) - \lim_{z \uparrow K_j} J(K_j)\ \ j = 1, 2$$

Finally, the amount of bunching of taxable incomes at K_2 amongst individuals with salary income s is

$$B_{K_2}(s) = J(K_2|s) - \lim_{z \uparrow K_2} J(K_2|s)$$

2.2.1. Special Case 1: Frictionless Benchmark. As a benchmark, this section studies a special case of the model in which all workers, indexed by i, are (i) unable to misreport their income: $\partial e(0,0)/\partial(s-\hat{s}) = \partial e(0,0)/\partial(n-\hat{n}) = \infty$; (ii) costlessly able to find their preferred job: $\phi_i = 0\,\forall i$; and (iii) fully sophisticated: $\lambda_i = 1\,\forall i$.

Workers' optimal salary and taxable incomes are given by

$$\{s_i^*, z_i^*\} = \begin{cases} \alpha_i p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}, (\alpha_i+\beta_i)\, p^{1+\varepsilon}(1-\tau_0)^{\varepsilon} & \text{if } \alpha_i+\beta_i < \underline{\delta}_1 \\ \frac{\alpha_i}{\alpha_i+\beta_i}K_1, K_1 & \text{if } \underline{\delta}_1 \le \alpha_i+\beta_i \le \bar{\delta}_1 \\ \alpha_i p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}, (\alpha_i+\beta_i)\, p^{1+\varepsilon}(1-\tau_1)^{\varepsilon} & \text{if } \bar{\delta}_1 < \alpha_i+\beta_i < \underline{\delta}_2 \\ \frac{\alpha_i}{\alpha_i+\beta_i}K_2, K_2 & \text{if } \underline{\delta}_2 \le \alpha_i+\beta_i \le \bar{\delta}_2 \\ \alpha_i p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}, (\alpha_i+\beta_i)\, p^{1+\varepsilon}(1-\tau_2)^{\varepsilon} & \text{if } \bar{\delta}_2 < \alpha_i+\beta_i \end{cases} \tag{2.2.2}$$

where $\underline{\delta}_j = K_j / \left[p^{1+\varepsilon}(1-\tau_{j-1})^{\varepsilon}\right]$ and $\bar{\delta}_j = K_j / \left[p^{1+\varepsilon}(1-\tau_j)^{\varepsilon}\right]$ for $j = 1,2$. Since the labour market is frictionless and misreporting is infinitely costly, these are also workers' equilibrium outcomes and reported incomes. The following lemmas summarize the predictions for bunching in the frictionless benchmark model.

LEMMA 1 (Frictionless Taxable Income Bunching). *The distribution of taxable incomes, $J^*(z)$ features excess bunching at the kinks K_1, K_2: $B_z(K_j) > 0$, $j = 1,2$*

PROOF. See appendix 2.7.1 □

Individuals' marginal incentive to accrue taxable income $1 - T'(z_i)$ jumps down as taxable incomes crosses a kink, so since the distribution of tastes $g(\alpha,\beta)$ is smooth, a mass of individuals choose to locate themselves at a kink. Turning to salary incomes,

LEMMA 2 (Frictionless Salary Bunching). *The distribution of the reported salary incomes for individuals with no non-salary income, $H^*(s^*|n^*=0)$ features excess bunching at the kinks K_1, K_2: $B_s(K_j|n^*>0) > 0$, $j = 1,2$. However, the distribution of the preferred salaries of individuals with non-salary income, $H^*(s^*|n^*>0)$ does not feature excess bunching at the kinks: $B_s(K_i|n^*=0) = 0$, $j = 1,2$.*

PROOF. See appendix 2.7.2 □

All individuals face incentives to have taxable incomes that bunch at the kinks. For individuals without non-salary income, this implies placing their salary income at a kink. However, for individuals with non-salary income, this is not the case. For these individuals, placing their taxable income at a kink implies placing their salary at a range of income levels in the interior of the tax brackets. Put differently, marginal incentives to accrue income $1 - T'(z_i)$ change as salary income crosses a kink for individuals without non-salary income, but not for individuals with non-salary income.

Finally, turning to how taxable income bunching varies with salary income,

LEMMA 3 (Frictionless TI Bunching Conditional on Salary). *The amount of excess bunching in taxable incomes at K_2 conditional on salary earnings s^*: $B^*_{K_2}(s^*)$ varies smoothly around K_1: $\lim_{s^* \uparrow K_1} B^*_{K_2}(s^*) = \lim_{s^* \downarrow K_1^*} B^*_{K_2}(s^*)$.*

PROOF. See appendix 2.7.3 □

As shown by equation (2.2.2), workers whose taxable income bunches at K_2 are those with $\underline{\delta}_2 \leq \alpha_i + \beta_i \leq \bar{\delta}_2$. The measure of this set of workers varies smoothly with α_i and hence with s_i^* since $g(\alpha, \beta)$ is smooth by assumption.

Having established these three properties of the frictionless equilibrium, I now turn to the equilibrium with frictions and study how these properties are affected by the presence of real adjustment costs and information frictions.

2.2.2. Special Case 2: Salary Misreporting. The first empirical predictions come from introducing the possibility of misreporting of incomes. I maintain the assumptions of costless labour market adjustment and full information, but allow individuals to misreport their incomes. Under the parameterization of evasion costs in 2.2.1, real decisions are undistorted by the possibility of evasion, so $s = s^*$, and $n = n^*$. Individuals must then choose whether to underreport their incomes and if so, by how much. If an individual misreports both her salary and non-salary income, her reports satisfy the first order conditions$e_s(s - \hat{s}^*, n - \hat{n}_s^*) = e_n(s - \hat{s}^*, n - \hat{n}_s^*) = \tau$, where e_s and e_n denote the partial derivatives of $e(s - \hat{s}, n - \hat{n})$ with respect to $s - \hat{s}$ and $n - \hat{n}$ respectively. By contrast, if she chooses only to misreport her non-salary income and avoid the fixed cost e_0 her choice of non-salary income report satisfies$e_n(0, n - \hat{n}_0^*) = \tau$. The worker then misreports her salary income iff

$$e_{0i} < \tau [s - \hat{s}^* + \hat{n}_0^* - \hat{n}_s^*] + e(0, n - \hat{n}_0^*) - e(s - \hat{s}^*, n - \hat{n}_s^*) = e_0^* \tag{2.2.3}$$

and so the fraction of workers who misreport their salary is $D(e_0^*)$. Intuitively, if misreporting salary income reduces the cost of misreporting non-salary income sufficiently (i.e. if e is sufficiently convex), then individuals will prefer to underreport both their salary and non-salary incomes rather than only non-salary income. That is, the cost savings from using a convex combination of salary and non-salary underreporting rather than only non-salary misreporting outweight the fixed cost of misreporting salary.

PREDICTION 1 (Misreporting and Marginal Tax Rates). *Individuals facing higher marginal tax rates τ, are more likely to misreport their salary: $de_0^*/d\tau > 0$. Those that misreport their salary also misreport it by more: $d(s - \hat{s})/d\tau > 0$*

PROOF. The first part follows immediately from application of the implicit function theorem to (2.2.3). The second part follows from inspection of the first order condition for $s - \hat{s}$. □

Intuitively, the bigger the marginal tax rate, the greater the returns to underrporting income, and so the more likely individuals are to be willing to do so. Furthermore, individuals with more non-salary income are more likely to misreport their salary:

PREDICTION 2 (Misreporting and Self-Reported Income). *Individuals with larger non-salary (self-reported) incomes are more likely to misreport their salary: $de_0^*/d\beta > 0$. Those that misreport their salary also misreport it by more: $d(s - \hat{s})/d\beta > 0$.*

PROOF. See appendix 2.7.4. □

Since individuals with higher non-salary incomes would like to underreport their incomes by more, they face a stronger incentive to also misreport their salary income, meaning that more people with higher non-salary incomes will also misreport their salaries. We can also summarize the implications of the presence of misreporting for bunching at the kinks as follows

PREDICTION 3 (Misreporting and Taxable Income Bunching). *Bunching of reported taxable incomes at kinks is stronger in the presence of evasion than without evasion:*

$$B_z\left(K_j|\hat{z}\leq z\right) > B_z\left(K_j|\hat{z}=z\right)$$

PROOF. See appendix 2.7.5. □

The ability to misreport incomes makes taxable income more responsive as individuals have an additional margin along which to adjust. As a result, reported income is more sensitive to the tax rate, and so bunching is stronger.

2.2.3. Special Case 3: Firm Responses to Adjustment Costs. The second set of empirical predictions comes from a special case focusing on the role of adjustment cost in the labour market. For this, I assume that (i) workers are unable to misreport their income $\partial e\left(0,0\right)/\partial\left(s-\hat{s}\right)=\partial e\left(0,0\right)/\partial\left(n-\hat{n}\right)=\infty$; (ii) a proportion δ of individuals faces no search costs ($\phi_i=0$) while the remaining workers have infinite search costs; and (iii) all workers are fully sophisticated: $\lambda_i=1\,\forall i$.

With these assumptions the labour market equilibrium is very simply characterized. Workers have preferred salaries chosen as in (2.2.2), giving rise to a distribution of preferred hours $F^*\left(l\right)$. Workers who face no adjustment costs choose their preferred salaries, and have hours distributed according to the aggregate distribution of worker preferences $F^*\left(l\right)$, while workers with adjustment costs have salaries distributed according to the offer distribution $F^o\left(h\right)$. Therefore, the search process maps the distribution of offers and the distribution of worker preferences to a distribution of accepted offers according to

$$\mathcal{D}\left[F^o\right]=\delta F^*\left(l\right)+\left(1-\delta\right)F^o\left(l\right) \tag{2.2.4}$$

In equilibrium, the distribution of offers must equal the distribution of accepted salaries (a fixed point of $\mathcal{D}$) which here means that the distribution of offers matches the aggregate distribution of workers' preferences, $F^o\left(l\right)=F^*\left(l\right)$. Since workers without non-salary income have salary preferences featuring bunching at the kinks (by lemma 2) the aggregate distribution of workers' preferences will also feature bunching at the kinks, and as a result the distribution of accepted salaries of *all* workers, including those with non-salary income, will feature this *firm bunching* at the kinks (Chetty *et al.*, 2011). In particular, PREDICTION 4 (Firm Bunching). *The equilibrium distributions of salaries for workers both with and without non-salary income features excess bunching at the kinks in the tax schedule due to firm offers featuring excess bunching at the kinks – Firm Bunching:* $B_s(K_j|\hat{n} > 0) > 0$ *and* $B_s(K_j|\hat{n} = 0) > 0$ $j = 1, 2$.

PROOF. The aggregate distribution of preferred hours is given by

$$F^*(l) = \eta F^*(l|n^* = 0) + (1 - \eta) F^*(l|n^* > 0)$$

which features bunching at the kinks since by lemma 2 $F^*(l|n^* = 0)$ features bunching at the kinks. For workers without non-salary income, the equilibrium distribution of offers is given by

$$\begin{aligned} F^e(l|n^* = 0) &= \delta F^*(l|n^* = 0) + (1 - \delta) F^*(l) \\ &= [\delta + (1 - \delta)\eta] F^*(l|n^* = 0) + (1 - \delta)(1 - \eta) F^*(l|n^* > 0) \end{aligned}$$

which again features bunching at the kinks due to the first term, coming from workers without non-salary income. Similarly for workers with non-salary income, the equilibrium distribution of accepted salaries is

$$\begin{aligned} F^e(l|n^* > 0) &= \delta F^*(l|n^* > 0) + (1 - \delta) F^*(l) \\ &= (1 - \delta)\eta F^*(l|n^* = 0) + [(1 - \delta)(1 - \eta) + \delta] F^*(l|n^* > 0) \end{aligned}$$

and this distribution also features bunching due to the preferences of workers without non-salary income. □

Despite the fact that workers with non-salary income do not face a kink in their budget set when their salary is at a kink in the tax schedule, the distribution of salaries they accept will feature bunching around the kinks. This bunching arises because firms, being unable to condition offers on worker characteristics such as sophistication and non-salary earnings, tailor offers to aggregate preferences, which feature bunching around the kinks. This firm bunching is absent in the frictionless model in section 2.2.1 because workers can costlessly adjust their salaries to find their preferred salary-hours packages (even though the overall distribution of salaries is the same). This means that looking for excess bunching in the distribution of salary incomes of workers with non-salary income provides a test for the presence of firm bunching, and consequently for the presence of fixed adjustment costs.

2.2.4. Special Case 4: Income Shifting Responses. Firm bunching attenuates the impact of adjustment costs by tailoring salary hours packages to the average preferences of workers. For workers with representative preferences – workers without non-salary income, this helps to mitigate the negative effects of the existence of adjustment costs. By contrast, for workers with unrepresentative preferences – workers *with* non-salary income, firm bunching makes it more difficult for workers to find a salary hours package fitting their preferences. However, these workers potentially have another means of responding - they can adjust the amount of their non-salary earnings in response to the salary income offers they receive.

This section characterizes the extent to which workers are able to respond by adjusting non-salary earnings. I focus on how much taxable incomes are able to respond to kinks in the tax schedule – the amount of bunching at K_2 by workers with nonsalary income – in the presence of firm bunching. Consider an equilibrium in which (i) workers are unable to misreport their income $\partial e\left(0,0\right)/\partial\left(s-\hat{s}\right)=\partial e\left(0,0\right)/\partial\left(n-\hat{n}\right)=\infty$; (ii) all workers have non-salary income ($\eta=0$); (iii) a proportion δ of individuals faces no search costs ($\phi_i=0$) while the remaining workers have infinite search costs; and (iv) all workers are fully sophisticated: $\lambda_i=1\,\forall i$. In this case,

PREDICTION 5 (Income Shifting and Double Bunching). *Workers respond to adjustment costs in salary earnings determination by shifting income between salary and non-salary earnings. The strength of this response is governed by the substitutability between l and q in the utility function, σ. In particular, we have that (i) $B_z\left(K_2|\phi_i=0\right)\geq B_z\left(K_2|\phi_i=\infty\right)$; (ii) $B_z\left(K_2|\phi_i=0\right)=B_z\left(K_2|\phi_i=\infty\right)\iff\sigma=1$; and (iii) $B_z\left(K_2|\phi_i=0\right)-B_z\left(K_2|\phi_i=\infty\right)$ is increasing in σ.*

PROOF. See appendix 2.7.6 □

Adjustment costs sometimes force workers to accept suboptimal jobs. If salary and non-salary earnings are perfect substitutes ($\sigma=1$), then this doesn't affect the worker's taxable income, as she simply shifts from salary to non-salary income. However, if she is unable to perfectly substitute between salary and non-salary income ($\sigma>1$), then being constrained in her choice of salary hours will impact on her taxable income, and make it more difficult for her to bunch her taxable income at a kink. As a result, if we still observe strong bunching of taxable incomes amongst individuals facing adjustment costs, this implies that salary and non-salary incomes are readily substitutable, so that σ is modest.

2.2.5. Special Case 5: Learning by Bunching. The final empirical prediction comes from introducing information frictions. I first abstract from adjustment costs and show that if receiving a salary at K_1 makes neighbouring kinks more salient, we should expect more bunching at K_2 from workers with $s = K_1$ than from workers with nearby salaries — an *information effect*. I then introduce adjustment costs and show that when firm bunching pushes individuals to accept salaries at kinks, this causes an additional *mismatch effect* and reduces taxable income bunching at K_2. Finally, I combine these two effects and characterize the total change in the amount of excess bunching at K_2 expected for individuals with salary $s = K_1$ compared to nearby salaries.

To see the information effect, consider an equilibrium without evasion or adjustment costs: $\partial e\left(0,0\right)/\partial\left(s-\hat{s}\right) = \partial e\left(0,0\right)/\partial\left(n-\hat{n}\right) = \infty$, and $\phi_i = 0\,\forall i$ However, some individuals do not perceive the kinks ($\lambda_i = 0$). I assume that before searching for a job, all workers are equally likely to have perceived the kinks, but that when workers receive a firm offer at a kink, this makes the kinks salient and this increases the probability that a worker is sophisticated by $\Delta\gamma$. Denoting the probability that a worker is sophisticated conditional on his/her salary as $\gamma\left(s\right) \equiv \mathsf{P}\left(\lambda = 1|s\right)$, this amounts to assuming that

$$\gamma\left(s\right) = \begin{cases} \bar{\gamma} & \text{if } s \notin \{K_1, K_2\} \\ \bar{\gamma} + \Delta\gamma & \text{if } s \in \{K_1, K_2\} \end{cases} \tag{2.2.5}$$

For my empirical strategy, what is important here is not that γ is constant away from the kinks, but that it is continuous everywhere except at K_1, where it jumps up due to the kinks becoming salient. Since naïve workers do not perceive the kinks in the tax schedule, I assume that amongst these workers there is no excess bunching of taxable incomes around K_2.

LEMMA 4 (Information Effect). *Excess bunching in taxable income as a function of salary income jumps up discretely at $s = K_1$.*

$$\begin{aligned} B_{K_2}(s) &= \gamma(s) B^*_{K_2}(s|\lambda_i = 1) \\ &= [\bar{\gamma} + \Delta\gamma I\{s = K_1\}] B^*_{K_2}(s|\lambda_i = 1) \end{aligned}$$

where $B^*_{K_2}(s)$ is the amount of taxable income bunching at K_2 by workers who receive salary s and accurately perceive the tax schedule.

PROOF. By assumption, the distribution of taxable incomes amongst individuals who do not perceive the tax schedule properly $(\lambda_i = 0)$ does not feature excess bunching around the kinks ($B_{K_2}(s|\lambda_i = 0) = 0$), so the amount of taxable income bunching at K_2 conditional on salary level s is $B_{K_2}(s) = \gamma(s) B_{K_2}(s|\lambda_i = 1) + [1 - \gamma(s)] \times 0$. The result then follows by noting that by lemma 3 $B_{K_2}(s|\lambda_i = 1)$ is smooth everywhere, and $\gamma(s)$ is smooth everywhere except at the kinks. □

To see the mismatch effect, reintroduce adjustment costs so that a proportion δ of workers have $\phi_i = 0$ while the remainder have $\phi_i = \infty$. Now consider individuals who accurately perceive the tax schedule $(\lambda_i = 1)$, for these workers we have,

LEMMA 5 (Mismatch Effect). *The probability that an individual with salary s has taxable income bunching at K_2 jumps down discretely at $s = K_1$. The amount of excess bunching at K_2 per worker with salary s is given by*

$$p(s) = \tilde{\delta}(s) B_{K_2}(s|\phi_i = 0) + \left(1 - \tilde{\delta}\right) B_{K_2}(s|\phi_i = \infty) \tag{2.2.6}$$

where

$$\tilde{\delta}(s) = \frac{\delta}{\delta + (1 - \delta)\left[f^o(s) / f^*(s|n > 0)\right]}$$

and $f^*(s|n \neq 0)$ is the density of preferred salary incomes for individuals with non-salary income, $f^o(s)$ is the density of firm salary offers.

PROOF. At any salary level there are $\delta f^*(s|n > 0)$ individuals with $\phi_i = 0$ who have chosen salary s because it is their preferred salary. $B_{K_2}(s|\phi_i = 0)$ of these individuals have a taxable income that bunches at K_2. The remaining $(1 - \delta) f^o(s)$ individuals have $\phi_i = \infty$ and have accepted a salary offer at s despite it not being their preferred salary.

$B_{K_2}(s|\phi_i = \infty)$ of these individuals bunch at K_2. Therefore, by Bayes' rule, the probability that an individual with salary s also has $\phi_i = 0$ is $\tilde{\delta}(s)$. By prediction 4 $f^o(s)$ features bunching at $s = K_1$, while by lemma 2 $f^*(s|n > 0)$ does not. Therefore, $\tilde{\delta}(s)$ jumps down discretely at $s = K_1$ and $p(s)$ assigns greater weight to bunching amongst constrained individuals. Finally, note that by prediction 5 $B_{K_2}(s|\phi_i = \infty) < B_{K_2}(s|\phi_i = 0)$ and the result follows. □

Having characterized the effects of both adjustment costs and information frictions separately, on taxable income bunching at K_2, I can now combine them to characterize the implications of both effects together on taxable income bunching at K_2.

PREDICTION 6 (Learning By Bunching). *The probability that an individual with salary income s has taxable income bunching at K_2 jumps discretely at $s = K_1$, and the proportional jump is*

$$\frac{p(K_1)}{\lim_{x\to K_1} p(x)} = \left(1 + \frac{\Delta\gamma}{\bar{\gamma}}\right)\frac{1 + b_S\tilde{B}}{1 + b_S} \tag{2.2.7}$$

where $b_S \equiv B_s(K_1) / \lim_{s\to K_1} h(s)$ is normalized excess firm bunching of salaries at K_1, and $\tilde{B} \equiv B_{K_2}(K_1|\phi_i = 0) / \left[\tilde{\delta}B_{K_2}(K_1|\phi_i = 0) + \left(1 - \tilde{\delta}\right)B_{K_2}(K_1|\phi_i = \infty)\right] \in [0, 1]$.

PROOF. The full proof is in appendix 2.7.7. Intuitively though, the first term captures the information effect as characterized in lemma 4, which increases taxable income bunching at K_1, while the second term captures the mismatch effect as characterized in lemma 5, which reduces taxable income bunching at K_1 □

Equation 2.2.7 characterizes the effect of adjustment costs and information frictions on workers' propensity to be taxable income bunchers as a function of their salary. In section 2.5 I will estimate $\frac{p(K_1)}{p(x)}$, while in section 2.4.3 I estimate firm bunching b_S. While this still leaves $\Delta\gamma/\bar{\gamma}$ underidentified, note that it can be bounded as

$$\frac{p(K_1)}{\lim_{x\to K_1} p(x)} - 1 \leq \frac{\Delta\gamma}{\bar{\gamma}} \leq \frac{p(K_1)}{\lim_{x\to K_1} p(x)}(1 + b_S) - 1 \tag{2.2.8}$$

since $\tilde{B}$ is bounded between 0 (mismatch effect completely eliminates TI bunching) and 1 (mismatch effect is 0).

2.3. Context & Data

Pakistan is a large developing country with a population of around 190 million. Tax revenues represent only 9% of GDP, a small amount even by lower income country (LIC) standards (Gordon & Li, 2009). Of total revenues, around 60% is collected through various withholding regimes, including the income tax on salaried employees that I study here, one of the bedrocks of the Pakistani tax system. For the perwonal income tax, an indvidual's taxable income is the sum of an individual's salary income, business income, capital income, foreign-source income and "other" income minus charitable deductions.[9] There are two tax schedules for the taxation of individual income, depending on the composition of an individual's taxable income. If an individual's salary income is less than half of their taxable income, they are taxed as self-employed individuals, using a tax schedule featuring notches (discrete jumps in the average tax rate) at threshold incomes (these notches form the basis of the paper by Kleven & Waseem, 2013). By contrast, in this paper I focus on individuals whose salary income represents more than half of their taxable income, who are taxed using the schedule for salaried employees (roughly a quarter of income tax returns are filed by individuals taxed as salaried employees). These individuals face a complicated tax schedule featuring between 17 and 20 income brackets (or "slabs" as they are known in Pakistan).

At the thresholds between brackets, the average tax rate on income jumps up creating a discrete jump up in the individual's tax liability—a notch. Panel A of figure 17 shows an example of a budget constraint in consumption-earnings space affected by a notch at K at which the average tax rate jumps from τ_1 up to τ_2. However, a complex system known as marginal relief was introduced in 2008 to smooth these notches, creating a pair of kinks at each bracket threshold, one convex and one non-convex. The marginal relief system allows taxpayers with incomes above the bracket threshold to opt to pay a high marginal tax rate τ_M on the income they earn above the threshold. Panel B of figure 17 shows the marginal relief schedule as the blue, dashed line in the budget set diagram. This smooths the discontinuous jump in tax liability at the notch, replacing it with a sharp convex kink where the marginal tax rate jumps from τ_1 to τ_M. At some point, however, the marginal relief no longer minimizes an individual's tax liability and she optimally opts to pay τ_2 on

[9] Agricultural income is taxed by the provincial governments separately

her entire income, creating a concave kink where the marginal tax rate jumps down from τ_M to τ_2. Panel C of figure 17 shows the tax-minimizing tax schedule around K. Table 5 shows the full schedule for the year 2009-10, giving a sense of how complicated the tax schedule is.

Two additional features of the Pakistani setting are important to note. First, as is common accross the world, employers of salaried workers are required to withold income tax on their employees (Slemrod, 2008) treating their salary as if it was their total taxable income, and to remit the tax to the government on their employees' behalf. In addition, employers are required to declare all their employees, their gross salaries, and the tax withheld on them to the tax authorities. However, apart from withholding income tax, firms have no other tax or benefit obligations linked to the level of the salaries they report. In particular, there is no payroll tax, and there are no social security contributions linked to workers' salaries.[10] This means that both firms and workers have incentives to underreport salaries, in contrast to the Mexican setting studied by Kumler *et al.* (2013) in which payroll taxes and benefits linked to reported salaries generate opposite incentives for firms and workers to misreport salaries. Furthermore, since employers are able to deduct their entire wage bill (salaried plus non-salaried employees) from their corporate income tax liability, underreporting salaries need not affect their corporate tax liability as it can be accompanied by overreporting non-salaried wages.

Second, trade unions that could be determining wages at a collective level are almost completely absent in Pakistan, particularly amongst the salaried, private sector workers I study. Fewer than 3% of formal sector workers in Pakistan are unionized, only 1% are under collective bargaining agreements, and unions mostly represent public sector workers (Mehmood, 2012), while the data used here covers only the private sector. The two biggest unions in Pakistan are the railway workers' union and the airline union, and all results are robust to excluding them. In addition, a number of textile firms have unionized workers, but these unions mostly represent contract workers, not salaried workers, and so do not appear in my data.

[10] Some employers do make pension contributions linked to workers' pay, but these contributions are not reported to FBR, and are, in any case independent of the salary reported to FBR.

2.3.1. Data. This paper is one of the first to have access to administrative data on the universe of taxpayers in a low income country.[11] I use data from income tax returns from Pakistan covering the fiscal years 2007/08–2011/12 (though I focus mostly on 2008/09–2011/12 as this is the period during which the kinked schedule described above is in place)[12] from the Federal Board of Revenue, Pakistan (FBR). I also use third-party reports on salaries from Employer Statements in which employers declare their employees' salaries and income tax witheld (the equivalent of the W-2 form in the United States).

I merge the employer statements with the income tax returns to have both salary and taxable income data for workers, and both employer and employee reports of the workers' salary. As shown in table 10, which outlines the merging procedure, the match rate is just over 50%. This rate is pulled down by two factors. First, the Employer Statement data only covers the private sector, and a large part of the salaried workforce is employed in the public sector. Second, the employer statements are not automatically checked for internal consistency, and so many records have missing or inaccurate identifiers, preventing a match with the income tax returns.

The salary data features strong bunching at round-number multiples of Rs. 5,000 in monthly terms (Rs. 60,000 in annual terms). In order to avoid conflating this heuristic bunching at round numbers with responses to kinks in the tax schedule, I drop the roughly 7.5% of jobs with round number salaries. Since 2 of the kinks in the tax schedule—at Rs. 900,000 and at Rs. 1,200,000—are at round numbers, I also exlude these two kinks from my analysis, though the results are robust to including them. After these steps, the dataset consists of 314,994 employee-employer-year observations.

The main variables I use are a worker's salary from the employer statements, and the worker's salary income, total income and taxable income (total income minus deductions) from the tax returns. In addition, I use a number of observable characteristics of the firms and workers as control variables and to estimate heterogeneity in bunching. Table 6 shows summary statistics of the matched dataset, and the subsample of workers whose taxable income differs from their salary (as reported by their employer) by at least 2%.

[11] predecessors include Kleven & Waseem, 2013; Kumler *et al.*, 2013; and Best *et al.*, 2013.

[12] The fiscal year in Pakistan runs from July 1 to June 30.

2.4. Taxable Income Responses: Kinks, Imperfect Enforcement, and Adjustment Costs

2.4.1. Taxable Income Responses: Bunching Around Kinks. The model in section 2.2 predicts that taxable incomes z will bunch around the kinks in the tax schedule. Conceptually, these taxable income responses combine real changes in earning behaviour by firms and workers; shifting responses as workers shift earnings across tax bases and/or between salary and non-salary earnings; and evasion responses. The following sections provide evidence on the presence of all three classes of responses, but I begin by establishing that there is clear evidence of sharp behavioural responses to the tax schedule by studying bunching of overall taxable incomes.

Figure 18 shows the distribution of taxable incomes around kinks in the tax schedule amongst all workers in the merged sample. Each individual's taxable income is scaled by the kink it is closest to, permitting me to pool all the kinks and years into a single figure.[13] The blue dots show the observed distribution of scaled taxable incomes, while the red line is the estimated counterfactual distribution in the absence of kinks in the tax schedule.[14] The figure also shows an estimate of the excess bunching mass in the distribution normalized by the counterfactual density at the kink, b: a statistic which permits comparison across figures, and which is proportional to the magnitude of the earnings response to the tax rate (Saez, 2010).[15] Bunching is significant and extremely sharp, $b = 1.08\ (0.127)$ demonstrating clearly that behaviour is responding to the tax schedule.[16]

---

[13]Bunching around the kinks is stable across the years in the sample, and as shown below, tends to be stronger at the lower kinks in the tax schedule, though it is present at all kinks.

[14]Grouping the data into bins 0.1% wide, the counterfactual is estimated as

$$c_j = \sum_{m=0}^{q} \beta_m \left(d_j\right)^m + \sum_{r=\bar{k}^-}^{\bar{k}^+} \gamma_r \mathbf{1}\left\{j = r\right\} + \mu_j \tag{2.4.1}$$

where c_j is the number of observations in bin j, d_j is the distance of bin j from a kink, $(j - 100)/0.1$, and q is the order of the polynomial ($q = 7$ in figure 18). The second term excludes bins in a region $\left[\bar{k}^-, \bar{k}^+\right]$ around the kinks, and μ_j is an error reflecting misspecification of the estimating equation. Standard errors are obtained by bootstrapping as in Chetty *et al.* (2011).

[15]Total excess mass around the kinks in the distribution is given by $B = \sum_{r=\bar{k}^-}^{\bar{k}^+} \left(c_r - \hat{c}_r\right)$ where $\hat{c}_r$ is the counterfactual mass in bin r predicted by estimating (2.4.1) omitting the contribution of the dummies for the excluded range around the kink, $\hat{c}_r = \sum_{m=0}^{q} \hat{\beta}_m \left(d_r\right)^m$. To permit comparison across distributions, the figures show $b = B/c_0$, the excess mass normalized by the average counterfactual density in the excluded range, $c_0 = \left[\frac{\bar{k}^+ - \bar{k}^-}{0.1}\right]^{-1} \sum_{r=\bar{k}^-}^{\bar{k}^+} \hat{c}_r$.

[16]This finding is estimated amongst workers who can be matched to their employer's salary report. Taxable income bunching is stronger in the full population of income tax filers $b = 1.66\ (0.115)$. However, this is mostly driven by the rate at which workers can be matched to their employers being higher at the higher kinks. Reweighting the income tax filers to match the composition of the merged sample, bunching becomes statistically indistinguishable in the two samples $b = 1.32\ (0.095)$.

While previous findings from high income countries have tended to find very diffuse bunching around kinks (Saez, 2010; Chetty *et al.*, 2011, 2013), the sharp bunching found here suggests that behavioural responses are strong and precise. This is despite the fact that this is a group of workers who we expect to have trouble adjusting reported earnings to taxes as they are likely to face rigid hours constraints (Rosen, 1976; Altonji & Paxson, 1988) and search costs (Rogerson *et al.*, 2005; Manning, 2011) in determining their salary earnings, which make it difficult for workers to target their earnings at kinks precisely. This is also among the first compellingly identified evidence of behavioural responses to individual income taxation from a developing country context. Kleven & Waseem (2013) provide evidence of behavioural responses from Pakistan, focusing mostly on an earlier time period during which the notched tax schedule was in place, and Kumler *et al.* (2013) study employer compliance with payroll taxation in Mexico.

2.4.2. Evasion Responses: Unilateral Salary Misreporting. The model presented in section 2.2 predicts that all workers will misreport their non-salary income, and that some workers will choose to misreport their salary income (predictions 1 & 2). In order to asess these predictions, I exploit the fact that I have independent reports of a worker's salary from the employee and the employer, and look for discrepancies between the two.

Despite employers withholding income tax on their employees, workers may wish to underreport their salary for a number of reasons. Most importantly, workers who also report non-salary income must remit the difference between the tax on their total taxable income and the tax that their employer has already withheld on their salary, so underreporting their salary reduces the amount they must pay. Even workers who only report salary income have incentives to underreport their salary. All individuals face a number of direct taxes on consumption items (effectively excise taxes) that are classed as income taxes for administrative purposes and reported together with the tax on taxable income, and so underreporting their salary will reduce their net tax liability. They may also wish to claim that their employers have overwithheld income tax on them and claim a tax refund.

Firms also have incentives to underreport salaries. Underreporting a salary reduces the amount that the employer has to withhold and remit to the government, relieving liquidity constraints and allowing employers to offer higher net of reported tax wages to their employees. Since both employers and employees have incentives to underreport salaries, they may collude in what they report. In this case, the two reports will be the same, but will both be smaller than the true salary. Since I am only focusing on unilateral underreporting by workers, the misreporting reported here should be interpreted as a very loose lower bound on the true extent of salary misreporting.[17]

Table 9 shows the extent of salary misreporting. 19.3% of workers unilaterally underreport their salary.[18] Furthermore, these workers underreport by a significant amount. Taking the firms' reports as the truth, workers who underreport understate their income by 15.6% leading to 3.6% of overall income going untaxed. The tax losses are larger still. To estimate the amount of tax lost, I calaculate the tax liability implied by each report, assuming that the reported salary is the worker's total taxable income. Since most workers do not have any non-salary income, this assumption is accurate in most cases. However, this assumption will tend to underestimate the effect of underreporting salary on the tax liability for workers with non-salary income due to the convexity of the tax schedule. As shown in panel C, evaders understate their tax liability by at least 21.3%, leading to a loss of at least 5.1% of tax revenue.

Figure 19 tests whether predictions 1 and 2 are borne out in the data. As predicted by the model, panel A shows clearly that individuals facing higher marginal tax rates are indeed more likely to underreport their salary with individuals facing the highest marginal rates almost 5 times more likely to underreport their salary. Panel B shows that there is also evidence that individuals facing higher marginal tax rates misreport their salary by more, though this effect is mainly concentrated in individuals in the upper tax brackets above 10%. Furthermore, consistent with prediction 2, panel C shows that individuals with a higher share of self-reported income are more likely to misreport their salary income,

[17] In principle, firms also have incentives to unilaterally underreport salaries. A particularly stark example is given in Yaniv (1988) who studies employers' incentives to underwithhold by reporting a lower salary to the tax authorities than to the worker, allowing firms to withhold more income tax on the worker than they remit to the authorities. However, workers only report their total salary earnings on their tax returns. This means that in cases where the employer reports less than the worker, I cannot distinguish between workers with a job that is not in the employer reports and firms that are underreporting, so this data does not permit an analysis of firm underrporting. I drop all workers that I observe in more than one employer statement, dropping around 9% of observations in the process.

[18] To avoid conflating underreporting with marginal differences due to rounding errors, I restrict attention to discrepancies of at least 0.25%

and panel D shows that individuals with greater non-salary income also report larger discrepancies, though this relationship is not statistically significant.

Prediction 3 predicts that taxable income bunching should be stronger amongst individuals with more evasion opportunities. Since direct measures of evasion opportunities are not available, I rely on proxies for evasion opportunities that the previous literature has identified.[19] Table 8 show the results of estimating taxable income bunching separately in various sub-samples. Bunching is significantly stronger around the lower kinks in the tax schedule, for workers employed by individually owned firms (as opposed to incorporated businesses or partnerships), for firms that are not registered for the Value Added Tax (VAT) and firms that are not under the purview of a Large Taxpayers Unit (LTU). Bunching is also stronger for workers at smaller firms and firms in the trading, construction and services sectors. This evidence is consistent with a significant part of taxable income responses being driven by evasion, as previous work has suggested that evasion should be negatively correlated with firm size (Kleven *et al.*, 2009; Bigio & Zilberman, 2011), with the increased papertrail from being in the VAT net (Pomeranz, 2013), and with the increased scrutiny from a LTU (Almunia & Lopez-Rodriguez, 2013). It is also consistent with the patterns of heterogeneity in *corporate* income tax evasion in Pakistan shown in Best *et al.* (2013).

The evidence on unilateral misreporting presented here is is in sharp contrast to the (limited) available evidence from rich countries. For example, Kleven *et al.* (2011) find that only 1.3% of of workers in Denmark underreport their third-party reported personal income (their salary), and that the underreported income is only 0.2%. Similarly, IRS (2012) estimates that the net tax gap for salaries in the United States is under 1%. Since the evidence in Kleven *et al.* (2011) is based on pre-post audit comparisons, it also includes any underreporting in which firms and workers collude, and any jobs that are completely unreported that are detected by the auditors, which are not included in the estimates presented here.

Overall, this suggests that evasion of third-party reported salary income in Pakistan is orders of magnitude larger than in high income countries. What is more, the sample studied here – those for whom both firm and worker reports of salary are available – is

[19] In principle, salary misreporting is a direct measure of evasion opportunities, but since this directly affects the bunching behaviour, estimating taxable income bunching separately on misreporters vs. non-misreporters would not be a suitable test of prediction 3.

likely to be the most compliant segment of the workforce. Workers whose employers fail to report their salary do not face the possibility that their report is cross-checked with the employer's and so the risk of detection is smaller, presumably increasing the amount of evasion. Conversely, it suggests that greater use of cross-checking firm and worker reports can lead to large increases in compliance and revenues.

2.4.3. Real Responses and Adjustment Costs: Firm Bunching. As shown by prediction 4 of the model in section 2.2, in the presence of adjustment costs on the part of workers and hours constraints on the part of firms, firm salary-hours offers will bunch at kinks. In order to establish whether firms are driving any of the bunching of workers' *taxable* incomes, I focus on the subset of workers who report significant non-salary earnings (the 2% sample) and investigate bunching in their *salary* incomes. As the tax schedule is a function of taxable income (the sum of salary and non-salary earnings, net of deductions), these individuals do not face a kink in their budget set when their salary is at a kink in the tax schedule and so should not have salaries bunched around kinks in the absence of adjustment costs and hours constraints (lemma 2). Therefore, if we find bunching in their salary incomes it is direct evidence that bunching is being partly driven by firms.

Figure 20 shows the findings. It shows the distribution of salary incomes (scaled by their closest kink) for individuals in the 2% sample. There is clear, sharp bunching around kinks: the normalized excess mass is $b = 2.14\ (0.211)$ indicating that firms are placing salary-hours offers around kinks, even for workers whom this does not benefit, consistent with the conceptual framework in section 2.2.[20] The only similar finding in previous work is Chetty *et al.* (2011), who find bunching of salaries at statutory kinks for workers with significant deductions in Denmark. However, collective wage bargaining is highly prevalent in Denmark, making it impossible to distinguish aggregation of worker preferences by firms or by trade unions. As discussed in section 2.3, the role of unions in Pakistan is insignificant, particularly for the salaried, private sector workers I am able to observe here.[21] The findings here are therefore the first to provide direct evidence of *firm* responses to worker incentives.

[20]The 2% sample uses a 2% threshold to classify individuals as having significant non-salary income. Table 11 shows that this result is robust to using alternative thresholds.

[21]Dropping workers in the two big unionized sectors - airline and railway workers does not change any of the findings.

A potential concern with the interpreting this finding as evidence that firms are responding to worker-level incentives is that workers with uncertain non-salary income may prefer to accept a salary near a kink before their uncertain non-salary income is realized so that their taxable income is near a kink in expectation. In this case, the uncertainty in these workers' non-salary income would cause their taxable incomes to bunch diffusely around kinks in the tax schedule. However, figure 21 shows that this is not the case. Taxable income bunching is stronger amongst workers with non-salary income than workers with only salary income ($b = 2.03\ (0.183)$ vs $b = 0.98\ (0.137)$). More importantly, taxable income bunching is just as sharp, suggesting that salary bunching is indeed driven by firms' offers.

Firm bunching is not uniform across firms and workers. Bunching of salary incomes is weaker amongst workers who only have salary income, $b = 1.02, (0.142)$ suggesting that it is particular types of firms that respond, and that they employ particular types of workers. Propensity score reweighting observations without non-salary income to account for all observable firm and worker characteristics (DiNardo *et al.*, 1996) raises the estimated degree of bunching in salaries to $b = 1.51\ (0.191)$, suggesting only half of the discrepancy between firm-worker pairs with non-salary income and without is accounted for by observable differences between these groups, with the remainder accounted for by unobservable characteristics of the firm and the worker such as the cost of misreporting salaries and firms' ability to substitute between labour and other inputs.[22]

2.4.4. Income Shifting Responses: Double Bunching. Prediction 5 of the model in section 2.2 is that taxable income bunching should be smaller amongst individuals who face large adjustment costs in their salary determination. Direct measures of adjustment costs are unavailable, but as shown in lemma 5, firm bunchers—individuals with salaries at a kink, are more likely to face large adjustment costs than other workers. This implies that double bunching—having salary bunching at one kink and taxable income bunching

[22]Workers are weighted by $\rho_i / (1 - \rho_i)$ where ρ_i is their estimated propensity to have non-salary income. The propensity scores are estimated using a year-specific cubic spline in salary income with knots at the kinks in the tax schedule; year-specific dummies for reporting other income sources (business, property, capital, other) or deductions (charitable, contributions to worker's welfare funds); year-industry dummies; year-region dummies; year-specific cubics in firm age, the firm's number of workers, firm sales, and the proportion of the firm's workers with salaries at a kink; cubics in firm age, worker age, and how long an individual has been registered for; dummies for gender, registration for VAT, and the tax year.

at another kink— should be small if workers are unable to shift income between salary and nonsalary earnings.

Figure 22 shows the distribution of taxable income for workers with salaries very near a kink (their employer's report of their salary is within 0.5% of a kink) but with taxable incomes away from that kink.[23] There is clearly strong and sharp bunching at the kinks—$b = 5.46\ (1.24)$. From this I can conclude that workers respond to adjustment costs in salary determination by adjusting their non-salary income, shifting earnings between salary and non-salary income. Furthermore, the strength of the bunching at the kinks suggests that this shifting is relatively easy for workers, i.e. that σ, the complementarity between salaried and nonsalaried hours worked, is modest.

Furthermore, this is not purely driven by workers misreporting their earnings. As figure 23 shows, there is still strong evidence of double-bunching amongst workers who don't underreport their salaries ($b = 6.69\ (1.81)$ in panel A) and who self-report a salary at a kink ($b = 6.28\ (1.38)$ in panel B). To see more clearly that workers are adjusting their non-salary earnings to the presence of adjustment costs, figure 24 shows part of the distribution of non-salary income for individuals with salaries within 0.5% of a kink (in blue circles), and for individuals with salaries more than 0.5% away from a kink, but within 2.5% of a kink (in orange diamonds). The distributions are clearly similar with the exception of the presence of clear bunching in the blue distribution at Rs. 50,000, Rs. 150,000, and to a lesser extent Rs. 250,000. A Kolmogorov-Smirnov test rejects the equality of the two distributions with $p = 0.0069$. These amounts coincide with the distances between various kinks in the tax schedule, confirming that individuals are adjusting their non-salary income.[24]

The shifting behaviour identified here is distinct from what the previous literature has usually focused on as it occurs within the personal income tax base, whereas previous work has tended to study shifting across bases (for example, between the corporate and personal income tax base) in response to differences in tax rates across bases (Gordon & Slemrod, 2000; Kleven & Schultz, 2013).[25] However, these spillovers from the taxation of

[23] I define a worker as having taxable income "away" from that kink if the closest kink to his/her salary income is not the same as the closest kink to his/her taxable income.

[24] These numbers are also salient, round numbers and individuals may have a tendency to report non-salary income at these round numbers, but this does not invalidate these findings as Rs. 50,000 is no rounder an amount for individuals with salaries at kinks than for individuals with salaries near kinks.

[25] Kleven & Waseem (2013) is a notable exception here that studies the same issue using different variation for identification. They study shifting between salary and non-salary income using the notch by the shift from the

salary income onto non-salary income are important for two reasons. First, individuals in lower income countries are much more likely to have both salary and non-salary income than people in high income countries, so these spillovers have a real relevance for taxable income elasticities and tax policy. Second, as shown in section 2.4.2, individuals with significant amounts of non-salary earnings are more likely to evade their tax liabilities by misreporting their salary income so shifting responses will affect the overall level of evasion of the income tax.

2.5. Learning by Bunching

This section presents results arguing that workers learn about the tax schedule through their interactions with employers to determine their salaries. In particular, receiving a salary at a kink teaches the workers about the importance of kinks, making them more likely to have taxable income bunching at a different kink. Section 2.5.1 takes a cross-sectional approach to demonstrate this, while section 2.5.2 uses an event study methodology to control for potential selection effects coming from assortative matching between sophisticated firms and workers.

2.5.1. Cross-Sectional Results. Consistent with prediction 6 of the model in section 2.2, this section shows that the probability that workers have taxable income bunching at a kink is discretely higher when their salary income is at a kink due to firm bunching, consistent with the presence of a large information effect. To do this, I first develop a methodology to estimate the counterfactual level of taxable income bunching that would have been observed were it not for firm bunching. To do this, I borrow from the bunching literature (Saez, 2010; Chetty *et al.*, 2011; Kleven & Waseem, 2013) and fit a flexible, high-order polynomial to binned data on the observed outcome using data near the kinks, but excluding data very near the kinks. I then use the predicted values from this estimate as my counterfactual at the kinks.

Under the identifying assumption that all firm and worker characteristics determining worker responsiveness covary smoothly with salary around the kinks, this method will identify a valid counterfactual for responsiveness in the absence of firm bunching. I pool

salaried tax schedule to the non-salaried tax schedule when salary income falls below half of taxable income described in section 2.3 whereas I study shifting between salary and non-salary income within the salaried tax schedule.

all the kinks together by scaling salaries by the closest kink to them, K_S. Then, within bins of scaled salaries, I calculate the fraction of individuals who have taxable incomes near (defined as being within 0.5%) a kink K_{TI}, but that kink is not the same kink that their salary is near, $K_S \neq K_{TI}$. I denote this conditional probability by p_j, the probability of taxable income bunching in a bin centered at j%. Grouping the data into bins of scaled salary 0.2% wide, I estimate the counterfactual conditional probability with the following polynomial:

$$p_j = \sum_{m=0}^{q} \beta_m \left(d_j\right)^m + \sum_{r=\bar{k}^-}^{\bar{k}^+} \gamma_r \mathbf{1}\left\{j = r\right\} + \mu_j \tag{2.5.1}$$

where d_j is the distance between bin j and a kink, $(j - 100)/0.2$, and q is the order of the polynomial ($q = 7$ in figures 25 & 26). The second term excludes bins in a region $[\bar{k}^-, \bar{k}^+]$ around the kinks reflecting the possibility that firm bunching may be slightly diffuse (though as shown in figure 20 firm bunching is extremely sharp). Finally, μ_j is a residual reflecting misspecification of the conditional probability equation (2.5.1).

From the estimates of the coefficients in equation (2.5.1), I calculate the counterfactual conditional probability in the excluded region $[\bar{k}^-, \bar{k}^+]$ as $\hat{p}_j = \sum_{m=0}^{q} \hat{\beta}_m \left(d_j\right)^m$, and my estimate of the discrete change in the conditional probability at the kinks is the proportional difference between the observed probability at the kink and the average counterfactual probability in the excluded region around the kink

$$\Delta p = \frac{p_{100}}{\left(\frac{0.2}{\bar{k}^+ - \bar{k}^- + 1}\right) \sum_{j=\bar{k}^-}^{\bar{k}^+} \hat{p}_j} - 1$$

I estimate the standard error of this estimate using by bootstrapping as in Chetty *et al.* (2011).

Figure 25 shows the baseline results. The blue circles show the observed conditional probabilities p_j, while the orange line is the estimated counterfactual conditional probability function estimated excluding bins ± 0.2% from a kink. The figure also shows the estimate of Δp, how much the probability changes at the kink due to firm bunching, along with its bootstrapped standard error. Figure 25 clearly shows a sharp spike in taxable income responsiveness for workers with salaries at a kink. The estimate of $\Delta p = 1.283$ (0.265) indicates that the responsiveness of workers affected by firm bunching is more than double that of workers with salaries near, but not at a kink. Using equation (2.2.8) and the estimate of firm bunching $b_S = 2.14\ (0.219)$, we can bound the information effect $\Delta\gamma/\bar{\gamma}$ as lying between $\Delta p = 1.283$ and $\Delta p \times (1 + b_S) = 4.029$, significantly larger than 0.

One potential concern with this could be that the findings are driven by misreporting of the worker's salary rather than responses by workers to their true salary. Figure 26 repeats the same exercise, but uses the worker's salary report instead of the employer's report. In this case, $\Delta p = 0.691\ (0.240)$, which is significantly smaller than the result in figure 25, and does not seem visually to be larger than the spikes at other salary levels due to noise. This implies that it is the true salary, not the salary that the worker (mis)reported that matters for the worker's information.

A more serious concern is with the identifying assumption that firms that offer salaries that bunch around the kinks are not differentially likely to employ workers who are more responsive to tax incentives, i.e. there is no assortative matching between firms and workers on tax responsiveness. To address this, I turn to an event study methodology to rule out time-invariant selection effects.

2.5.2. Event Study Results. This section pursues an event study methodology to allow me to control for fixed firm and worker characteristics such as sophistication and long-run responsiveness (Jacobson *et al.*, 1993; Hilger, 2013; Chetty *et al.*, 2013). The strategy consists of comparing the outcomes of individuals who experience a treatment event – receiving a salary at a kink in the tax schedule, to the outcomes of individuals who experience a control event – receiving a salary in the interior of a tax bracket.

Let $g \in \{K, I\}$ denote whether a worker experiences a salary at a kink, K, or in the interior, I in a year s. Let t denote the year an outcome y is observed, and define $q = t - s$ as event time, the number of years since the event g. The event study model for individual i's outcome y, allowing the effects of kink and interior events to vary by period is then

(2.5.2)

$$y_{i,g,t,s} = \alpha + \sum_{j=q^-}^{q^+} \mu_j^K \mathbf{1}\{g = K, q = j\} + \sum_{j=q^-}^{q^+} \mu_j^I \mathbf{1}\{g = I, q = j\} + \Gamma X_{i,g,t,s} + u_{i,g,t,s}$$

where $q^- < 0$ and $q^+ > 0$ are the minimum and maximum values of q, respectively, $X_{i,g,t,s}$ is a vector of observable covariates,[26] and $u_{i,g,t,s}$ is an independently distributed error term. The key advantages of the event study model over a standard difference in difference (DD) model are that it allows the effects to vary arbitrarily over time and does not impose a fixed difference between treatment and control groups. This flexible specification then permits assessment of whether the assumptions necessary for identification in a traditional DD framework are reasonable.

To operationalize the event study I define salary bunching as receiving a salary within 1% of a kink, and taxable income bunching as having taxable income within 1% of a kink.[27] For the event study I construct two samples. The "Kink" sample consists of all workers who receive a salary at a kink in year s and also receive a salary in the interior of a tax bracket in year $s-1$. The restriction on year $s-1$ salary is intended to facilitate the interpretation of the results as the impacts of first-time exposure to the kinks in the tax schedule rather than repeated exposure. The "Interior" sample consists of workers who receive a salary in the interior of a tax bracket in year s but work at a firm where at least 1 worker received a salary at a kink. These workers must also have received a salary in the interior in year $s-1$ to match the restriction on workers in the Kink sample. I also include data from 2007/08 to increase the time dimension of the panel.[28] Table 7 shows summary statistics of the Kinks and Interior samples.

The outcomes I analyze are a variety of bunching behaviours combining taxable income and salary income bunching as summary indicators of sophisticated tax responsiveness. I will be interested in both the contemporaneous impact β_0 and medium-run impact $\beta_{>0}$ of experiencing a salary at a kink rather than a salary in the interior of a tax bracket, which I will estimate as DD treatment effects from equations of the form

[26] $X_{i,g,t,s}$ contains fixed effects for event by observation years, firm sector, tax office, worker gender, having business income, property income, capital income, foreign income, other income, charitable deductions, and other deductions; year-specific cubics in firm size, firm age, the degree of firm bunching in the firm, worker age and time since worker registration for taxes; and a year-specific cubic spline in salary with knots at the kinks in the tax schedule.

[27] Results are similar though noisier when using 0.5% for taxable income bunching

[28] This data was not appropriate for use in the earlier analysis as it is a year in which the thresholds the workers faced were notches rather than kinks. See Kleven & Waseem (2013) for a detailed analysis of these notches.

$$\begin{aligned} y_{i,g,t,s} &= \delta+\lambda I\{g=K\}+\psi_{-1}I\{q=-1\}+\psi_0 I\{q=0\}+\psi_{>0}I\{q>0\} \\ &+\beta_{-1}I\{g=K,q=-1\}+\beta_0 I\{g=K,q=0\}+\beta_{>0}I\{g=K,q>0\} \\ &+\Gamma X_{i,g,t,s}+u_{i,g,t,s} \end{aligned} \tag{2.5.3}$$

where the ψ_{-1} and β_{-1} terms are included to account for the fact that as a result of the definition of the samples, salary bunching is mechanically 0 at event-time -1 in both samples.[29]

The identifying assumption required in order to interpret β_0 and $\beta_{>0}$ as the causal effects of receiving a salary at a kink is that

$$I\{g=K,q=0\},I\{g=K,q>0\}\perp u_{i,g,t_O,t_E}$$

which in economic terms requires that

(1) There be parallel trends between the treatment and control groups before the event is experienced: $\mu_j^K-\mu_j^I=\rho\,\forall j<0$
(2) Individuals do not anticipate receiving a salary at a kink and respond preemptively in period $q<0$
(3) There are no time-varying unobserved worker or firm characteristics that are correlated both with event time q and with responsiveness to taxes y.

I verify the reasonableness of condition 1 visually through inspection of the μ_j^G dummies for $j<0$. Condition 2 is unlikely to be a major concern as inflation is high and volatile over the period I study in Pakistan, making it difficult to predict future wages with the precision necessary to target the kinks. Any violation of condition 2 will likely increase responsiveness in years before the event however, attenuating my estimates. Condition 3 is addressed through the addition of a rich set of individual-year and firm-year controls, reducing the scope for unobserved factors to affect the estimates.

[29]Note that the DD estimators β_0 and $\beta_{>0}$ are related to the event study dummies μ_j^g according to $\beta_0=\left(\mu_0^K-\mu_0^I\right)-\left(-q^-\right)^{-1}\sum_{j=q^-}^{-1}\left(\mu_j^K-\mu_j^I\right)$ and $\beta_{>0}=\left(q^+\right)^{-1}\sum_{j=1}^{q^+}\left(\mu_j^K-\mu_j^I\right)-\left(-q^-\right)^{-1}\sum_{j=q^-}^{-1}\left(\mu_j^K-\mu_j^I\right)$.

Figure 27 shows the evolution of overall bunching in salary (in panel A) and taxable (in panel B) incomes in the Kink sample and the Interior sample. The blue circles show the estimated μ_j^K for the Kinks sample from equation (2.5.2), while the orange crosses show the μ_j^I from the Interior sample. Each panel also shows the estimated contemporaneous effect β_0 and medium-term effect $\beta_{>0}$ from estimating equation (2.5.3) alongside the mean pre-event level of bunching in the Kink sample. In both cases it is striking that the trends before the event in the Kinks and Interior samples are remarkably parallel. Furthermore, experiencing a salary at a kink has a large effect on future bunching behaviour. It increases future salary bunching by 1.7 percentage points, a 28% increase, and taxable income bunching by 0.9 percentage points, an increase of 36%.

Figure 28 shows event study results from decomposing taxable income bunching into its 3 constituent parts and is constructed in the same way as figure 27. It shows taxable income bunching rates when salary income bunches at a different kink (panel A), when salary income is not at a kink (panel B), and when salary income is at the same kink as taxable income (panel C). Though the results are considerably noisier as a result of the smaller samples, it is again striking that the pre trends in the two samples are remarkably parallel. Panels A and C show that there is a strong effect on taxable income bunching accompanied by salary bunching either at the same kink or at a neighbouring kink, with medium term bunching increasing by 100% for double bunchers in panel A, and 40% for those without non-salary income in panel C. However, panel B does not provide strong evidence of an effect on taxable income bunching when salary is not at a kink, suggesting that workers learn about the significance of the kinks in the tax schedule and seek out both salaries and taxable incomes at kinks in the medium term.

2.6. Conclusion

This paper has exploited unique access to employee-employer matched administrative tax data on firms and salaried workers in Pakistan to explore the underappreciated role of firms in determining how workers' taxable earnings respond to taxation. Consistent with the model presented in section 2.2, I present evidence on three ways in which firms affect workers' earnings responses. First, third-party reporting of salaries by employers makes underreporting taxable income more costly for workers. Second, firms' equilibrium salary-hours offers respond endogenously to the presence of adjustment costs in the labour market. Third, workers learn about the tax schedule from firms' salary offers, making them more responsive to taxation both in the same year and in subsequent years.

While third-party reporting of salaries raises the cost of misreporting, it has not eliminated misreporting, as 19% of workers still misreport their salaries. This casts doubt on the efficacy of the third-party reporting that recent work has suggested is central to tax enforcement (Kleven *et al.*, 2009; Pomeranz, 2013) in low-capacity environments where cross checks of multiple reports of the same tax base are absent or limited. Since salaried workers are generally the most compliant group of personal income tax payers, this suggests that the self-employed will be even more responsive to the tax schedule. Together these suggest that the returns to investment in fiscal capacity are large, particularly in cross-checking third-party reports of tax bases and increased scrutiny of individuals with non-salary income, both subjects for future work.

This paper has also shown that in addition to their central role in the collection of taxes, firms play a key role in mitigating the impact of adjustment costs and information frictions on workers' responsiveness to taxes. Firms reduce the impact of adjustment costs by aggregating the preferences of workers and this is manifested in bunching in their salary offers around kinks in the tax schedule. This equilibrium level response by firms to worker level incentives also indirectly increases the responsiveness of taxable earnings amongst individuals facing information frictions since firms respond to aggregate preferences, including those of workers who do not face information frictions.

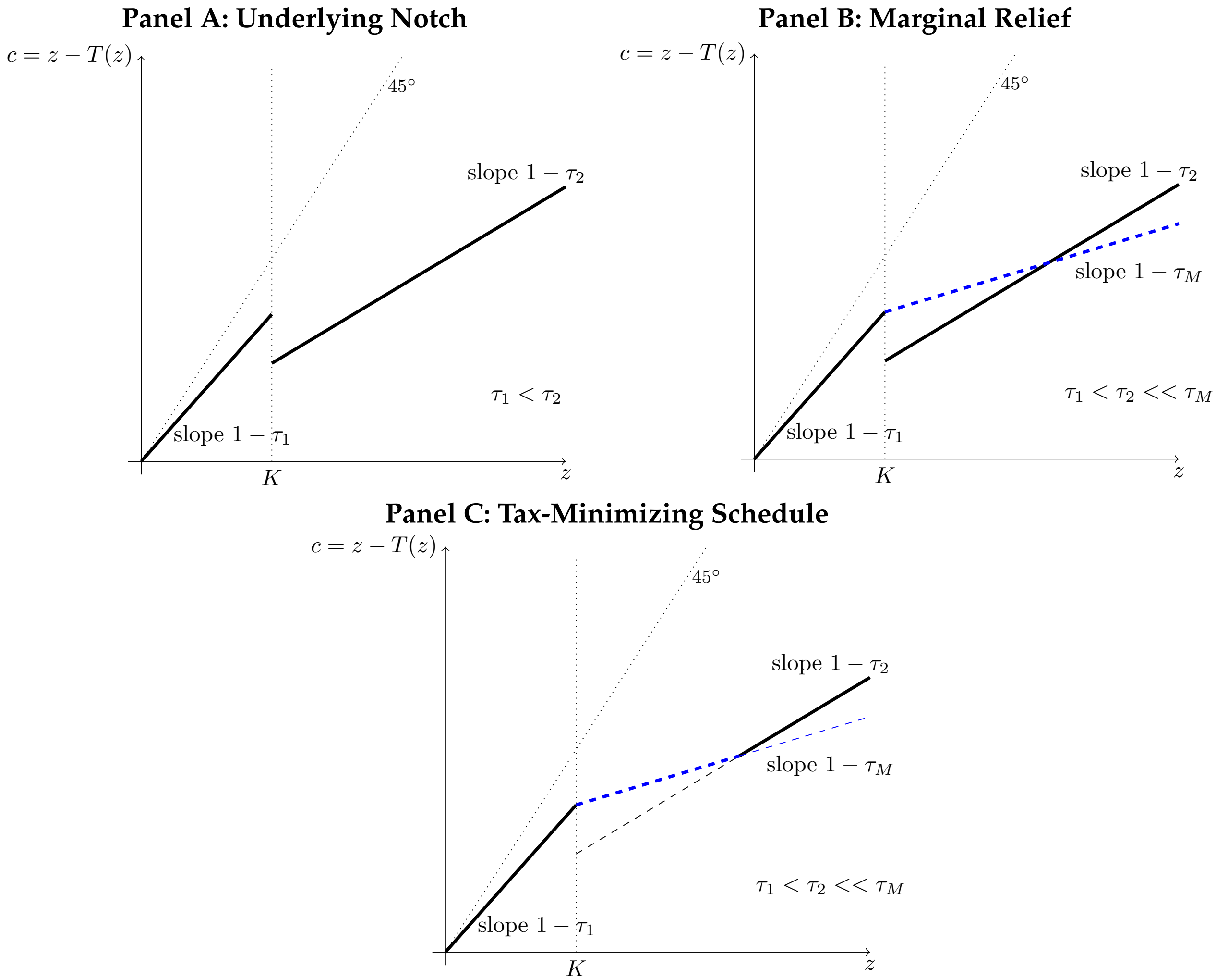
Furthermore, firm offers at kinks in the tax schedule directly affect the information frictions attenuating worker responses to the tax schedule. The effects of this are large, workers are around 130% more responsive to taxation in years they receive salaries at a kink, and 100% more responsive in future years, suggesting that information frictions play a large role in attenuating responses to the tax schedule. This implies that policies that make the tax schedule more salient, or simplify the tax schedule can have large impacts on how firms and workers respond to the tax schedule.[30] Of course, whether increasing responsiveness to the tax schedule improves welfare depends on whether individuals are

[30]Indeed, in 2012, the tax schedule was simplified to a set of 10 standard kinks, partly in response to the perception that the tax schedule was overly complicated. Of course, 10 kinks is still more than are found in most OECD countries suggesting further scope for simplification.

suffering a utility loss from their attenuated responsiveness due to optimization frictions (see Liebman & Zeckhauser, 2004; Bernheim & Rangel, 2009 for e.g.).

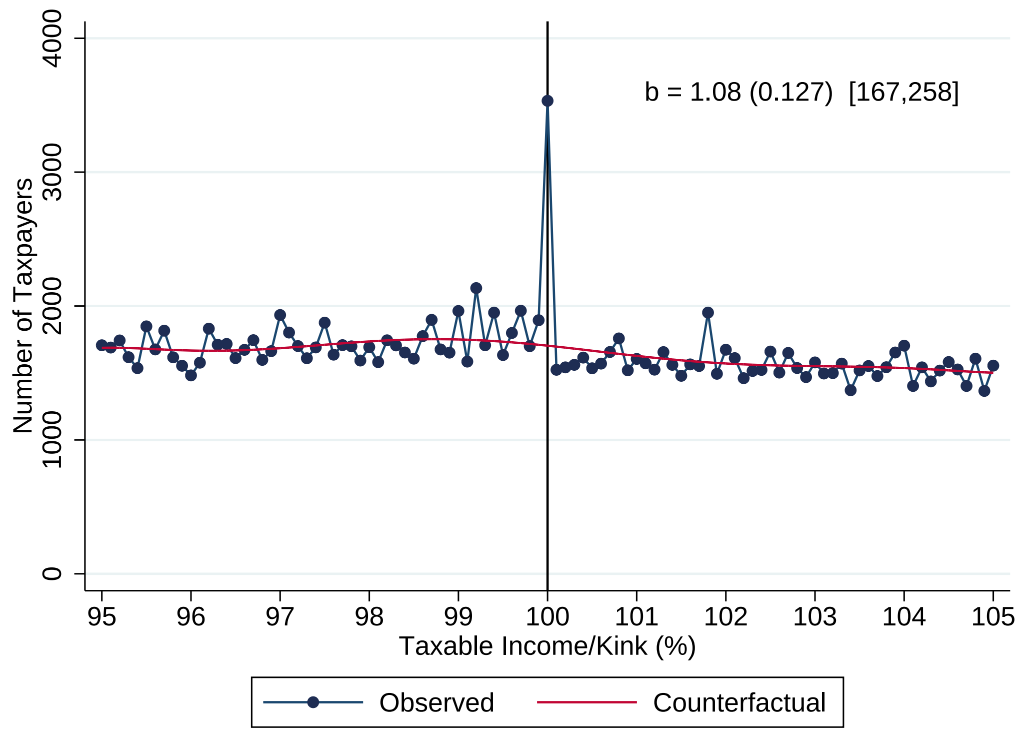
Overall, this paper has shown that in lower income country contexts firms affect evasion decisions, real earnings decisions, the impact of adjustment costs, and the information workers use in their decisions. In light of this, the virtual absence of firms in the public finance literature on income taxation (Kopczuk & Slemrod, 2006) has come at a great cost. Firms must play a central role in our analysis of income taxation in lower income countries.

FIGURE 17. EFFECT OF TAX SCHEDULE ON BUDGET SET



Notes: The figure shows the effect of the tax schedule for salaried workers $T(z)$ on a worker's budget constraint around a bracket threshold K. The budget constraint shows the relationship between consumption $c = z - T(z)$ and taxable income z. Panel A shows the underlying notch at the threshold K where the average tax rate jumps up from τ_1 to τ_2 generating a discrete fall in consumption at the threshold. Panel B shows the effect of marginal relief (the blue, dashed line) allowing taxpayers to opt to pay a high marginal tax rate $\tau_M >> \tau_2$ on their income above the threshold generating a convex kink at K where the marginal tax rate jumps up from τ_1 to τ_M. At some point, the marginal relief ceases to minimize the taxpayer's tax liability, and so an optimizing taxpayer opts to pay the flat rate τ_2 on their entire income, generating a concave kink where the marginal tax rate jumps down from τ_M to τ_2. Panel C shows the tax minimizing schedule around the threshold K combining the convex and concave kinks.

FIGURE 18. TAXABLE INCOME BUNCHING

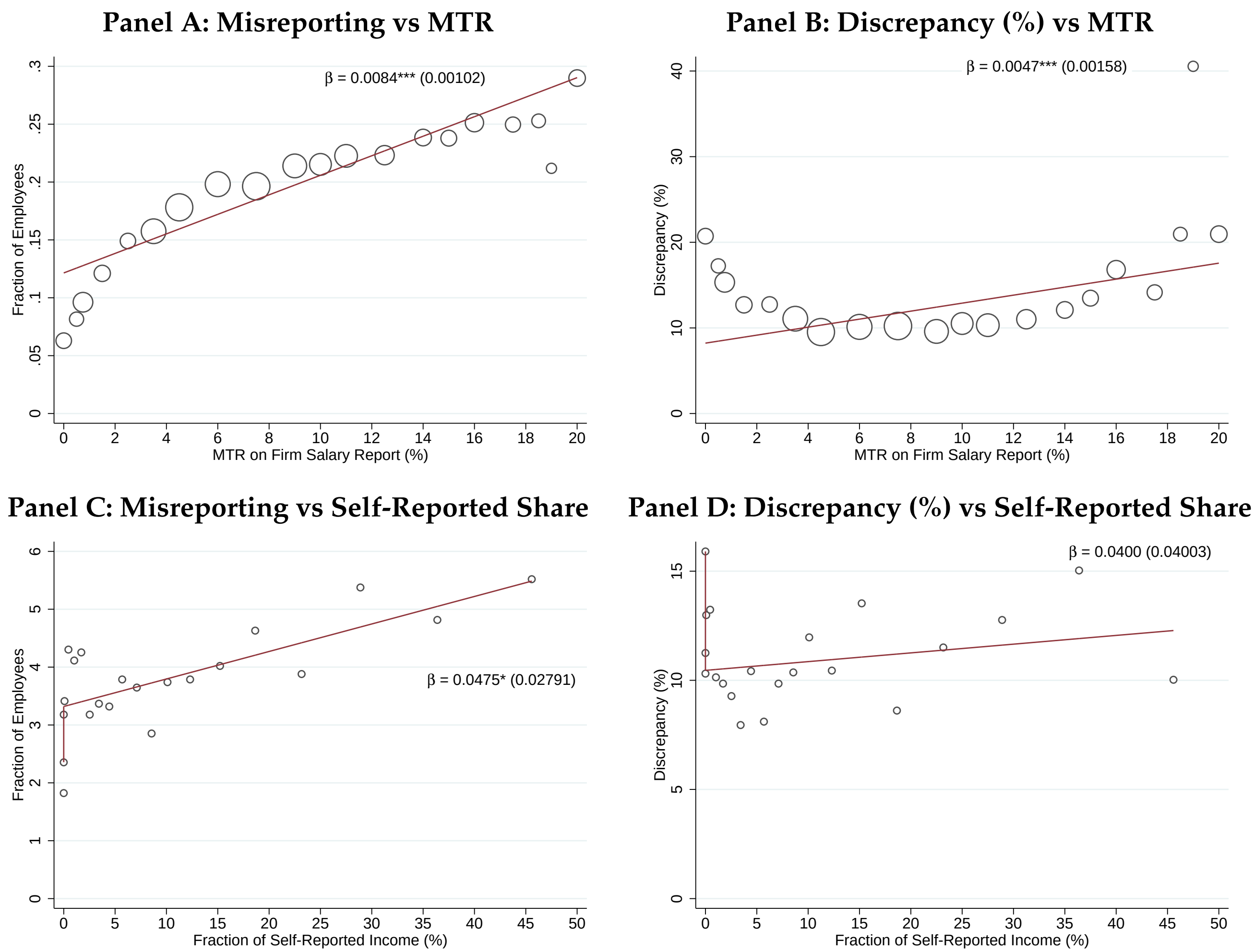


Notes: The figure shows the observed distribution of workers' taxable incomes (as a percentage of the nearest kink) in blue dots, alongside an estimate of the counterfactual density that would be observed if the tax schedule did not feature a kink at 100%. Grouping the data into bins 0.1% wide, the counterfactual is estimated as

$$c_j = \sum_{m=0}^{q} \beta_m (d_j)^m + \sum_{r=\bar{k}^-}^{\bar{k}^+} \gamma_r \mathbf{1}\{j = r\} + \mu_j \tag{2.6.1}$$

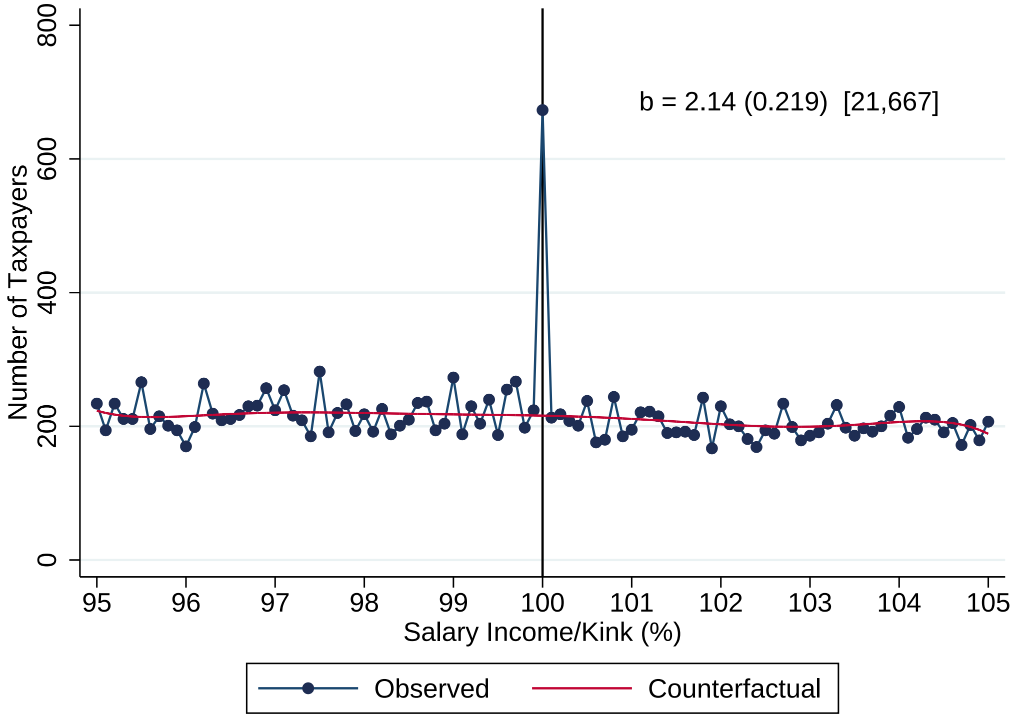
where c_j is the number of observations in bin j, d_j is the distance of bin j from a kink, $(j - 100)/0.1$, and q is the order of the polynomial ($q = 7$ in figure 18). The second term excludes bins in a region $[\bar{k}^-, \bar{k}^+]$ around the kinks, and μ_j is an error reflecting misspecification of the estimating equation. The figure also shows the normalized estimated excess mass in the observed distribution around the kinks. Total excess mass around the kinks in the distribution is given by $B = \sum_{r=\bar{k}^-}^{\bar{k}^+} (c_r - \hat{c}_r)$ where $\hat{c}_r$ is the counterfactual mass in bin r predicted by estimating (2.4.1) omitting the contribution of the dummies for the excluded range around the kink, $\hat{c}_r = \sum_{m=0}^{q} \hat{\beta}_m (d_r)^m$. The figure shows $b = B/c_0$, the excess mass normalized by the average counterfactual density in the excluded range, $c_0 = \left[\frac{\bar{k}^+ - \bar{k}^-}{0.1}\right]^{-1} \sum_{r=\bar{k}^-}^{\bar{k}^+} \hat{c}_r$. Standard errors are obtained by bootstrapping as in Chetty *et al.*, 2011. The number of observations used for the estimation is shown in square brackets.

FIGURE 19. CORRELATES OF MISREPORTING: PREDICTIONS 1 AND 2



Notes: Panels A and B show the correlation of the probability that a worker underreports his/her salary, and the size of the discrepancy amongst misreporters (as a percentage of the firm's salary report), with the marginal tax rate the worker faces, respectively. Panels C and D show the correlation between the probability that a worker underreports his/her salary, and the size of the discrepancy for misreporters and the marginal tax rate, and the share of his/her income that is self-reported, respectively. The fraction of self-reported income is calculated as the worker's reported non-salary income divided by the sum of the worker's reported non-salary income and the worker's employer's report of his/her salary. This measure is capped at 50% as above this the worker is no longer taxed as a salaried worker. In Panels A and B the grey circles show the averages within each tax rate, with the size of the circle proportional to the number of individuals facing each tax rate. The red line shows the fitted relationship from a linear OLS regression. In Panels C and D the grey circles show the averages within each vingtile of the distribution of the self-reported share. The red line shows the fitted

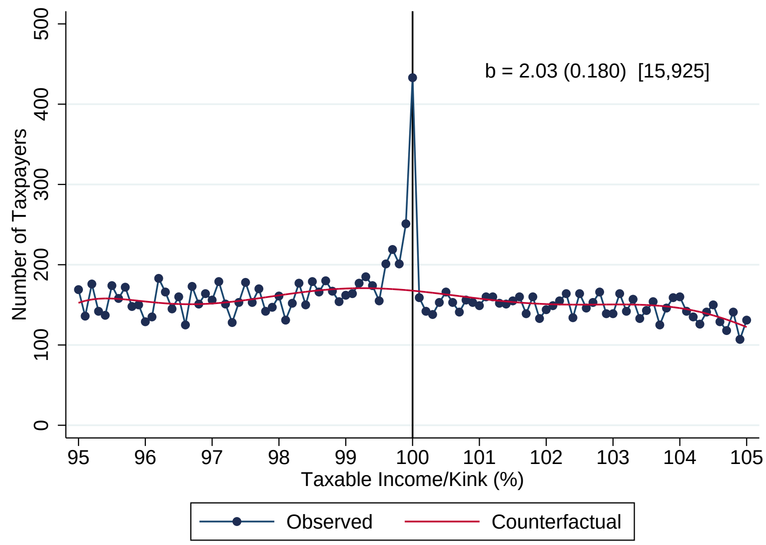
FIGURE 20. FIRM BUNCHING OF SALARIES



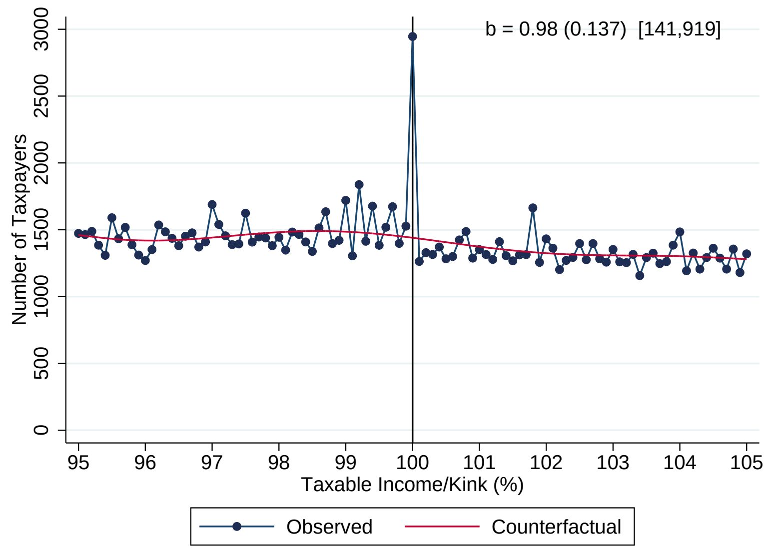
Notes: The figure shows the observed distribution of scaled salary incomes for workers with significant non-salary income (defined as having taxable income more than 2% different from salary income) in blue dots, alongside the estimated counterfactual distribution (red line) and the estimated normalized excess bunching mass b, its standard error in parentheses, and the number of observations used in square brackets (see notes to figure 18 for estimation details).

FIGURE 21. FIRM BUNCHING: FIRM RESPONSES OR UNCERTAIN INCOME?

Panel A: Taxable Income Distribution; Workers with Non-salary Income

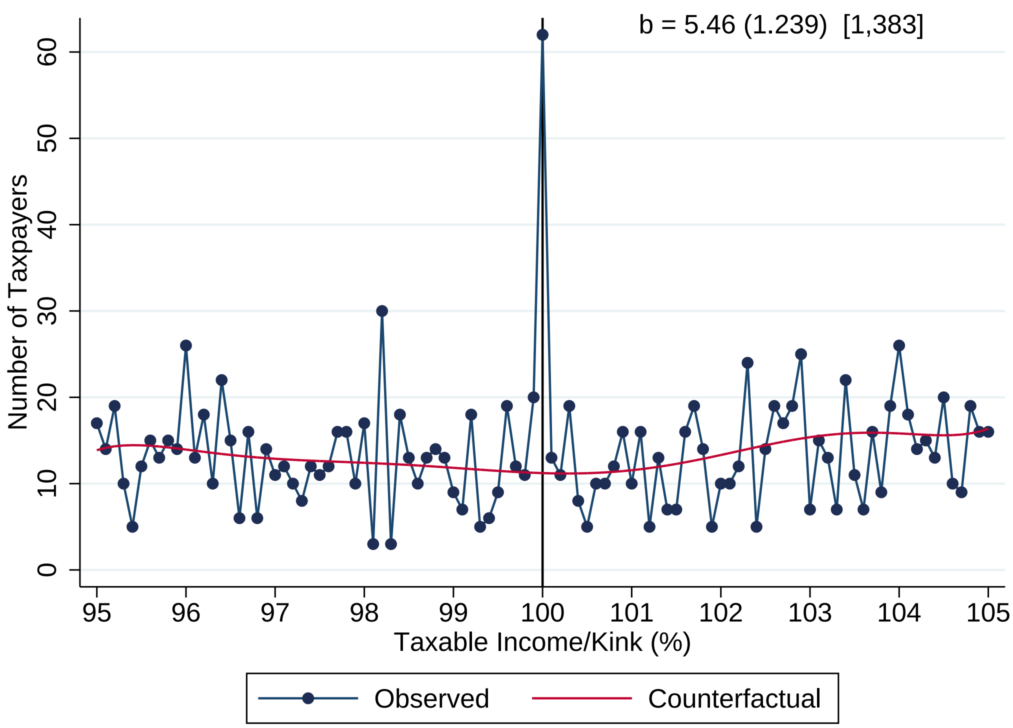


Panel B: Taxable Income Distribution; Workers without Non-salary Income



The figure shows bunching of scaled taxable incomes amongst workers with non-salary income (defined as having taxable income more than 2% different from salary income) in panel A, and without non-salary income in panel B. The figure shows the observed distributions in blue dots, alongside the estimated counterfactual distribution (red line) and the estimated normalized excess bunching mass b, its standard error in parentheses, and the number of observations used in square brackets (see notes to figure 18 for estimation details).

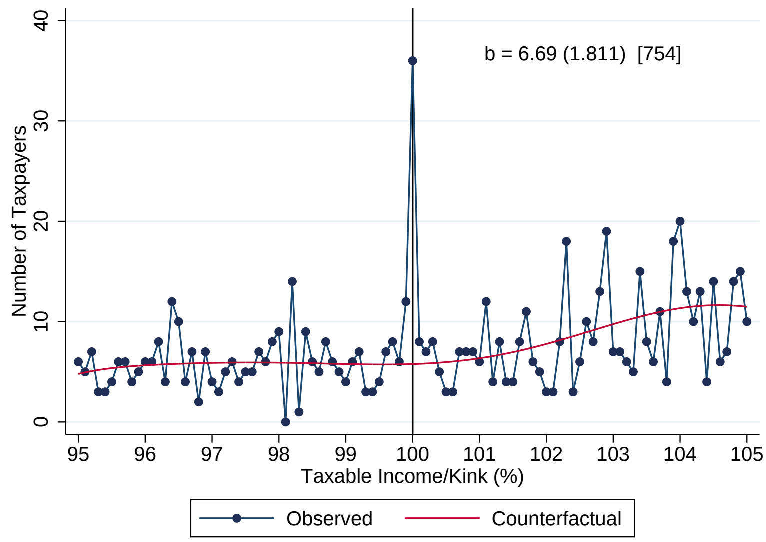
FIGURE 22. DOUBLE BUNCHING: TAXABLE INCOME DISTRIBUTION OF SALARY INCOME BUNCHERS



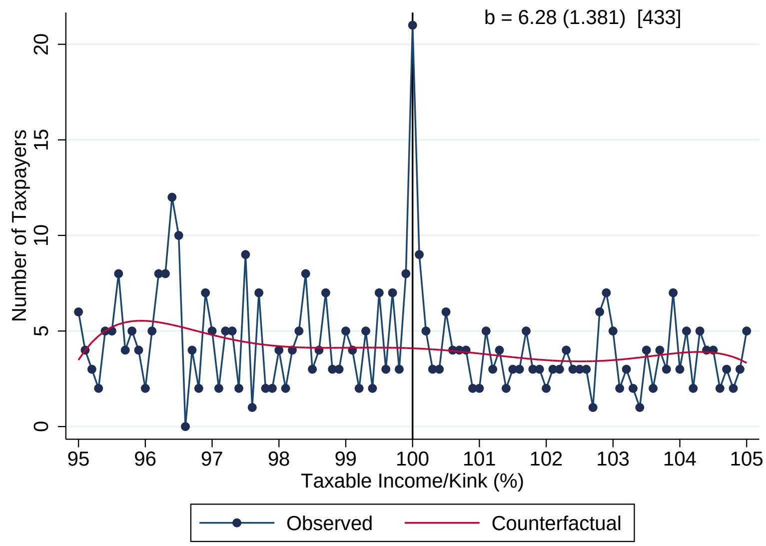
Notes: The figure shows taxable income distribution of workers with salary incomes near kinks, but with taxable incomes away from that kink. Workers are defined as having salary incomes near a kink if their employer reports a salary within 0.5% of a kink. Workers are defined as having taxable income away from that kink if the closest kink to their taxable income is not the same as the closest kink to their salary income. The blue dots show the observed distribution in 0.1% bins of scaled income, while the red line shows the estimated counterfactual distribution, and the panels also show estimates of the normalized excess bunching mass *b*. See the notes to figure 18 for details of the estimation methodology.

FIGURE 23. DOUBLE BUNCHING IS NOT DRIVEN BY SALARY MISREPORTING

Panel A: Employer-Reported SI at Kink, Employer Report $\geq$ Employee Report



Panel B: Employee-Reported SI at Kink

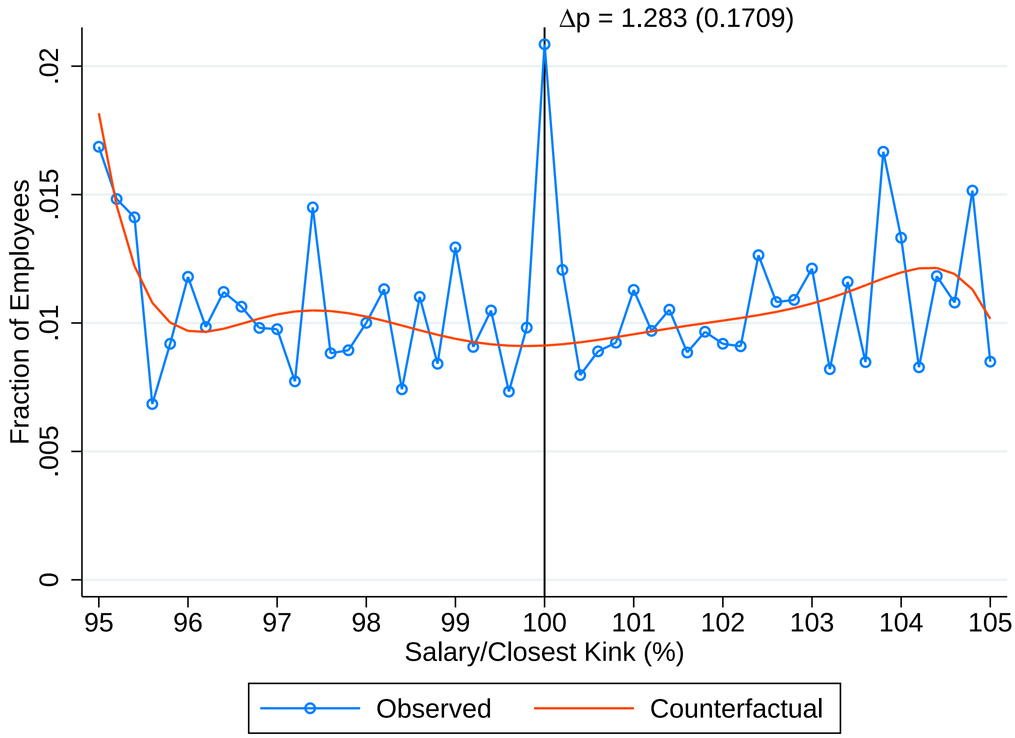


Notes: The figure shows taxable income distributions of workers with salary incomes near kinks. Workers are defined as having salary incomes near a kink if their employer reports a salary within 0.5% of a kink in panel A, or if the employee reports a salary within 0.5% of a kink in panel B. The panels show the distributions of taxable incomes for workers whose taxable income is away from the kink their salary is near. Specifically, if the closest kink to their taxable income is not the same as the closest kink to their salary income. The blue dots show the observed distribution in 0.1% bins of scaled income, while the red line shows the estimated counterfactual distribution, and the panels also show estimates of the normalized excess bunching mass b. See the notes to figure 18 for details of the estimation methodology.

FIGURE 24. DISTRIBUTIONS OF NON-SALARY INCOME

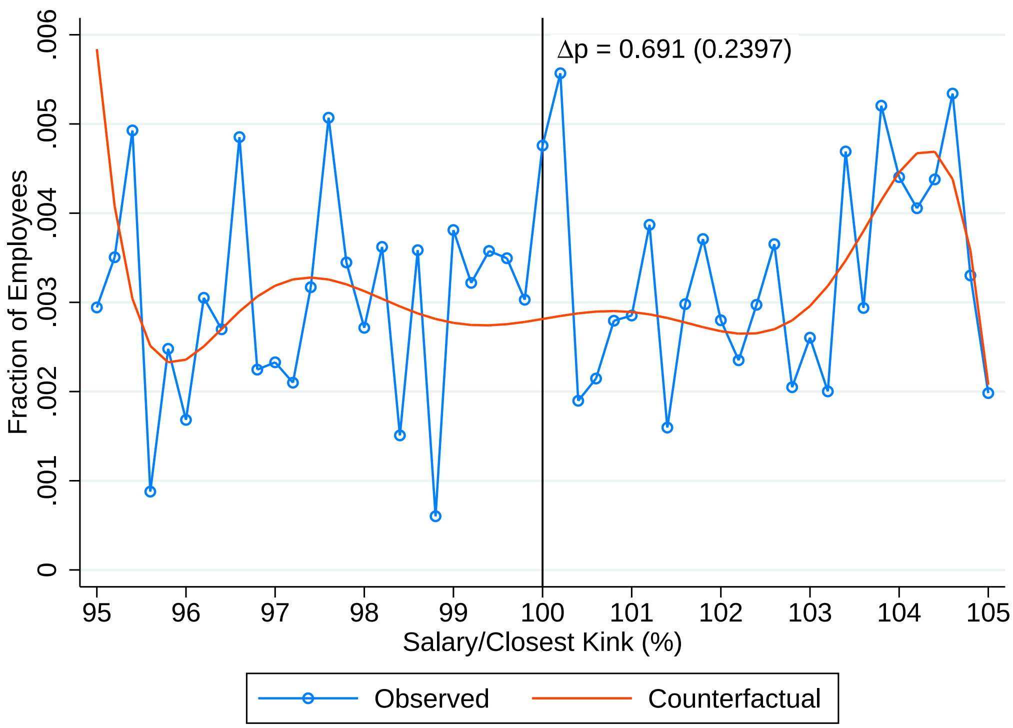
Notes: The figure shows the distributions of non-salary income (defined as the difference between taxable income and salary income) amongst workers with salaries at a kink (defined as being within 0.5% of a kink) in blue circles and workers with salaries near, but not at kinks (defined as being within 2.5% of a kink, but not within 0.5%) in orange diamonds. The figure also shows the p-value from a Kolmogorov-Smirnov test of the equality of the two distributions.

FIGURE 25. PROBABLITY OF TAXABLE INCOME BUNCHING AS DISTANCE OF SALARY FROM KINK VARIES



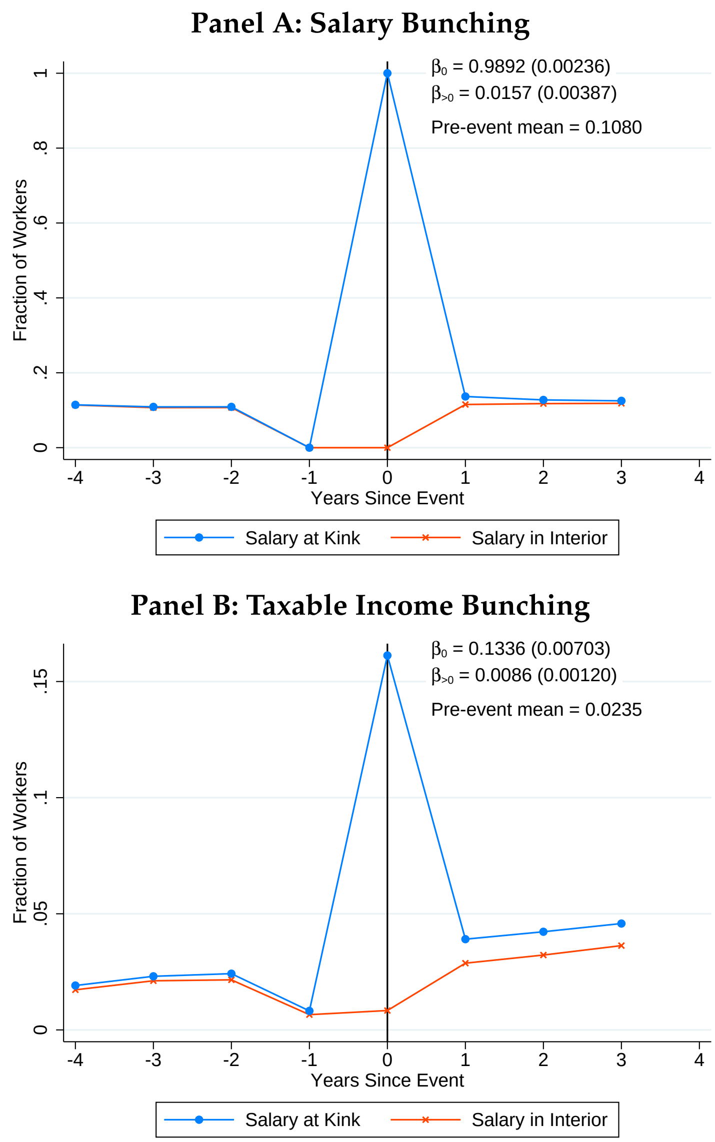
Notes: The figure shows how the probability that a worker has taxable income near a kink (defined as being within 0.5% of a kink) and that that kink is not the same kink as the closest kink to their salary, $K_{Si} \neq K_{TIi}$, changes as the distance between an individual's salary (as reported by his/her employer) and a kink varies. Each blue circle is the probability that a worker has taxable income near a kink within a bin of width 0.2%. The orange line is the estimated counterfactual probability estimated on the binned data using a 7th order polynomial as in equation (2.5.1) and excluding points in bins 0.2% above and below the kink. The figure also shows Δp, the observed increase in probability at the kink normalized by the average counterfactual probability in the excluded region around the kink, along with its standard error calculated by bootstrapping the procedure 200 times in brackets.

FIGURE 26. INFORMATION EFFECT IS NOT DRIVEN BY SALARY MISREPORTING



Notes: The figure shows how the probability that a worker has taxable income near a kink (defined as being within 0.5% of a kink) and that that kink is not the same kink as the closest kink to their salary, $K_{Si} \neq K_{TIi}$, changes as the distance between an individual's salary (as reported by the employee) and a kink varies. Each blue circle is the probability that a worker has taxable income near a kink within a bin of width 0.2%. The orange line is the estimated counterfactual probability estimated on the binned data using a 7th order polynomial as in equation (2.5.1) and excluding points in bins 0.2% above and below the kink. The figure also shows Δp, the observed increase in probability at the kink normalized by the average counterfactual probability in the excluded region around the kink, along with its standard error calculated by bootstrapping the procedure 200 times in brackets.

FIGURE 27. EVENT STUDY OF RECEIVING SALARY AT A KINK: OVERALL SALARY AND TAXABLE INCOME BUNCHING



Notes: The figure shows the evolution of bunching behaviour in the Kinks Sample, who experience a salary at a kink in year 0 (defined as a salary within 1% of a kink), and in the Interior Sample, who experience a salary in the interior of a tax bracket in year 0. Panel A shows salary bunching in the two samples. By definition of the samples salary bunching is 0 in both samples in year -1, 0 in the Interior sample in year 0, and 1 in the Kink sample in year 0. Panel B shows taxable income bunching in the two samples defined as having taxable income within 1% of a kink. The panels also show the estimated contemporaneous β_0 and medium term $\beta_{>0}$ effects of receiving a salary at a kink estimated from equation (2.5.3):

$$\begin{aligned} y_{i,g,t,s} \quad = \quad & \delta + \lambda \mathbf{1}\left\{g = K\right\} + \psi_{-1}\mathbf{1}\left\{q = -1\right\} + \psi_0 \mathbf{1}\left\{q = 0\right\} + \psi_{>0}\mathbf{1}\left\{q > 0\right\} \\ & + \beta_{-1}\mathbf{1}\left\{g = K, q = -1\right\} + \beta_0 \mathbf{1}\left\{g = K, q = 0\right\} + \beta_{>0}\mathbf{1}\left\{g = K, q > 0\right\} + \Gamma X_{i,g,t,s} + u_{i,g,t,s} \end{aligned}$$

The standard errors shown are robust standard errors clustered at the g, t, s level. The figures also show the pre-event mean in the Kinks Sample.

FIGURE 28. EVENT STUDY OF RECEIVING SALARY AT A KINK: DECOMPOSITION OF TAXABLE INCOME BUNCHING

Notes: The figure shows the evolution of the 3 components of taxable income bunching (defined as having taxable income within 1% of a kink) behaviour in the Kinks Sample, who experience a salary at a kink in year 0 (defined as a salary within 1% of a kink), and in the Interior Sample, who experience a salary in the interior of a tax bracket in year 0. The panels show the probabilities that taxable incomes and salaries bunch at different kinks (panel A) that taxable incomes bunch while salaries do not (panel B) and that taxable incomes and salaries bunch at the same kink (panel C) in the two samples. By definition of the samples bunching is 0 in both samples in year -1, and 0 in the Interior sample in year 0 in panels A and C; and 0 in the Kinks sample in panel B. The panels also show the estimated contemporaneous β_0 and medium term $\beta_{>0}$ effects of receiving a salary at a kink estimated from equation (2.5.3):

$$y_{i,g,t,s} = \delta + \lambda \mathbf{1}\{g = K\} + \psi_{-1}\mathbf{1}\{q = -1\} + \psi_0 \mathbf{1}\{q = 0\} + \psi_{>0}\mathbf{1}\{q > 0\} + \beta_{-1}\mathbf{1}\{g = K, q = -1\} + \beta_0 \mathbf{1}\{g = K, q = 0\} + \beta_{>0}\mathbf{1}\{g = K, q > 0\} + \Gamma X_{i,g,t,s} + u_{i,g,t,s}$$

TABLE 5. TAX SCHEDULE FOR SALARIED EMPLOYEES IN TAX YEAR 2009/10

From (Rs. 000s)	To (Rs. 000s)	Flat Rate (%)	Marginal Relief Rate (%)
0	200	0	
200	250	0.5	20
250	350	0.75	20
350	400	1.5	20
400	450	2.5	20
450	550	3.5	20
550	650	4.5	30
650	750	6	30
750	900	7.5	30
900	1,050	9	30
1,050	1,200	10	40
1,200	1,450	11	40
1,450	1,700	12.5	40
1,700	1,950	14	40
1,950	2,250	15	40
2,250	2,850	16	50
2,850	3,550	17.5	50
3,550	4,550	18.5	50
4,550	8,650	19	60
8,650	∞	20	60

Notes: The table shows the tax schedule for salaried employees in the tax year from 1 July 2009 to 30 June 2010. Each row represents a bracket of the tax schedule with its lower and upper bounds in the first two columns. The third column shows the flat average tax rate within the bracket, and the fourth column shows the marginal rate at which individuals can opt to be taxed on their income above the lower bound of the tax bracket.

TABLE 6. SUMMARY STATISTICS

Variable	Matched Sample Mean	s.d.	2% Sample Mean	s.d.
Salary (Employer Report)	1,460,681	3,579,009	1,456,672	4,737,086
I{Salary ≈ Kink}	0.061	0.239	0.056	0.231
Salary (Employee Report)	1,712,665	73,329,352	2,139,826	119,638,823
I{Salary ≈ Kink}	0.059	0.236	0.055	0.227
Taxable Income	1,882,843	159,527,501	2,582,172	260,363,676
I{TI ≈ Kink}	0.060	0.238	0.055	0.228
Total Income	1,851,126	159,181,413	2,532,930	259,796,228
I{Business Income}	0.031	0.173	0.077	0.266
I{Capital Income}	0.001	0.036	0.003	0.051
I{Foreign Income}	0.000	0.015	0.000	0.022
I{Other Income}	0.014	0.118	0.029	0.168
I{Deductions}	0.063	0.244	0.095	0.293
I{Zakat Deductions}	0.047	0.212	0.072	0.259
I{WWF Deductions}	0.005	0.070	0.008	0.089
I{Charitable Deductions}	0.020	0.141	0.031	0.173
Age	43.0	11.75	43.1	12.46
I{Female}	0.035	0.183	0.037	0.189
Years Reg. For Tax	9.0	5.39	9.1	5.36
I{Reg for VAT}	0.049	0.216	0.093	0.291
Firm ♯ of Workers	1751.9	3295.37	1651.0	3225.64
Firm Sales (Rs. Millions)	13,202.82	60,794.638	12,845.64	64,760.363
Firm Salary Bunching	0.066	0.072	0.068	0.076
Firm Age	11.0	4.73	10.5	4.94
I{Firm Reg for VAT}	0.798	0.402	0.763	0.425
I{Firm Under LTU}	0.632	0.482	0.577	0.494
I{Corporate Employer}	0.943	0.232	0.936	0.245
I{Individual Employer}	0.013	0.113	0.014	0.116
I{Partnership Employer}	0.044	0.205	0.051	0.219
I{Agriculture}	0.011	0.105	0.007	0.085
I{Construction}	0.015	0.123	0.015	0.122
I{Finance}	0.177	0.381	0.174	0.379
I{Manufacturing}	0.338	0.473	0.298	0.457
I{Mining}	0.038	0.192	0.038	0.192
I{Services}	0.353	0.478	0.402	0.490
I{Trading}	0.024	0.153	0.026	0.160
I{Utilities}	0.033	0.179	0.028	0.165
I{Other}	0.010	0.099	0.011	0.102
2008/09 (♯ of obs)	78,070		26,671	
2009/10 (♯ of obs)	81,536		26,594	
2010/11 (♯ of obs)	74,254		29,055	
2011/12 (♯ of obs)	81,134		35,912	
Overall (♯ of obs)	314,994		118,232	

Notes: The table shows means and standard deviations of variables in the matched sample and the 2% sample whose taxable income differs from their employer-reported salary by more than 2%. Income being "≈ Kink" is defined as being within 0.5% of a kink. Zakat deductions are religious charitable giving, collected centrally by the state in Pakistan. WWF Deductions are employers' tax-deductible contributions to a workers' welfare fund. VAT is the Value Added Tax (called the generalised sales tax in Pakistan), and the LTU is the Large Taxpayers Unit.

TABLE 7. SUMMARY STATISTICS OF KINKS AND INTERIOR SAMPLES

Variable	Kinks Sample		Interior Sample	
	Mean	s.d.	Mean	s.d.
Salary (Employer Report)	617,137	3,058,758	619,935	1,443,355
I{Salary $\approx$ Kink}	0.306	0.461	0.053	0.224
I{TI $\approx$ Kink}	0.056	0.230	0.015	0.120
I{Business Income}	0.006	0.076	0.006	0.075
I{Capital Income}	0.000	0.015	0.000	0.016
I{Foreign Income}	0.000	0.008	0.000	0.007
I{Other Income}	0.003	0.050	0.002	0.049
I{Deductions}	0.076	0.265	0.077	0.267
I{Zakat Deductions}	0.007	0.085	0.007	0.084
I{WWF Deductions}	0.001	0.029	0.001	0.028
I{Charitable Deductions}	0.003	0.052	0.003	0.055
Age	9.5	18.33	8.7	17.70
I{Female}	0.030	0.170	0.033	0.178
Years Reg. For Tax	6.8	5.53	6.3	5.54
Firm ♯ of Workers	2961.8	4235.82	2808.2	4108.55
Firm Sales (Rs. Millions)	20,019.76	67,312.533	18,540.23	62,783.602
Firm Salary Bunching	0.069	0.070	0.064	0.044
Firm Age	11.0	4.45	11.0	4.44
I{Firm Under LTU}	0.702	0.457	0.719	0.449
I{Corporate Employer}	0.964	0.186	0.969	0.173
I{Individual Employer}	0.008	0.091	0.006	0.078
I{Partnership Employer}	0.027	0.163	0.024	0.155
I{Agriculture}	0.010	0.097	0.010	0.098
I{Construction}	0.014	0.118	0.014	0.116
I{Finance}	0.186	0.389	0.185	0.388
I{Manufacturing}	0.337	0.473	0.340	0.474
I{Mining}	0.060	0.237	0.059	0.235
I{Services}	0.346	0.476	0.349	0.477
I{Trading}	0.018	0.134	0.016	0.124
I{Utilities}	0.026	0.158	0.024	0.153
I{Other}	0.003	0.059	0.003	0.058
2007/08 (♯ of obs)	61,789		145,053	
2008/09 (♯ of obs)	90,756		221,942	
2009/10 (♯ of obs)	98,581		244,071	
2010/11 (♯ of obs)	93,526		228,142	
2011/12 (♯ of obs)	87,715		209,264	
Overall (♯ of obs)	432,367		1,048,472	
Overall (♯ of workers)	101,758		355,158	

Notes: The table shows means and standard deviations of variables in the Kinks Sample, who experience a salary at a kink in year 0 (defined as a salary within 1% of a kink), and in the Interior Sample, who experience a salary in the interior of a tax bracket in year 0. Income being "$\approx$ Kink" is defined as being within 1% of a kink. Zakat deductions are religious charitable giving, collected centrally by the state in Pakistan. WWF Deductions are employers' tax-deductible contributions to a workers' welfare fund. The LTU is the Large Taxpayers Unit.

TABLE 8. TAXABLE INCOME BUNCHING AND PROXIES FOR EVASION OPPORTUNITIES

Worker Characteristics		Firm Characteristics				Sector	
TI ≤ Median	1.76 (0.205) [93,812]	**Corporate Employer**	1.30 (0.116) [156,957]	**# of Workers ≤ Median**	2.44 (0.237) [83,577]	**Agriculture**	0.70 (0.502) [1,927]
TI > Median	0.75 (0.092) [73,254]	**Individual Employer**	3.06 (0.642) [2,225]	**# of Workers > Median**	0.21 (0.069) [82,525]	**Mining**	0.25 (0.220) [5,685]
Age ≤ Median	1.28 (0.141) [57,570]	**Partnership Employer**	1.96 (0.326) [7,458]	**Sales ≤ Median**	1.81 (0.171) [83,876]	**Construction**	1.75 (0.472) [2,550]
Age > Median	1.38 (0.138) [52,697]	**Years Registered ≤ Median**	1.83 (0.169) [89,093]	**Sales > Median**	0.84 (0.096) [82,226]	**Manufacturing**	1.54 (0.176) [57,076]
Male	1.33 (0.123) [157,727]	**Years Registered > Median**	0.80 (0.100) [76,920]	**Size ≤ Median**	2.32 (0.218) [83,839]	**Trading**	2.65 (0.459) [3,985]
Female	2.63 (0.373) [5,884]	**Not VAT Registered**	1.87 (0.179) [33,180]	**Size > Median**	0.32 (0.073) [82,263]	**Services**	1.75 (0.162) [58,822]
Years Registered ≤ Median	1.33 (0.146) [84,987]	**VAT Registered**	1.20 (0.121) [130,827]			**Finance**	0.40 (0.115) [29,166]
Years Registered > Median	1.38 (0.122) [82,067]	**Not LTU**	3.07 (0.314) [61,406]			**Utilities**	0.50 (0.291) [5,796]
Not VAT Registered	1.17 (0.116) [75,801]	**LTU**	0.34 (0.065) [105,248]				
VAT Registered	5.58 (0.746) [2,587]						

Notes: The table shows estimated bunching of taxable income in various subsamples. For each subsample, the table shows the estimated normalized excess bunching mass b estimated as in section 2.4.1; the standard error of the estimate in round brackets, and the number of observations used for the estimation (those within 5% of a kink) in square brackets. A firm's (relative) size combines its sales and its number of employees by defining its size as the sum of its percentile in the distribution of number of workers and its percentile in the distribution of firm sales.

TABLE 9. PREVALENCE OF SALARY MISREPORTING

	Variable	Amount
Panel A: Underreporters (% of Workers)		
(1)	Employee < Employer	19.3
Panel B: Underreported Salary Income (SI)		
(2)	Employee < Employer (Rs. Bn)	15.6
(3)	Total Evaders' Employer Reported SI (Rs. Bn)	98.9
(4)	Total Employer Reported SI (Rs. Bn)	437.3
(5)	Employee Underreported SI (% of evaders' SI)	15.7
(6)	Employee Underreported SI (% of total SI)	3.6
Panel C: Underreported Tax Liability		
(7)	Employee < Employer (Rs. Bn)	3.1
(8)	Total Evaders' Employer Reported Tax (Rs. Bn)	14.4
(9)	Total Employer Reported Tax (Rs. Bn)	60.6
(10)	Employee Underreported Tax (% of evaders' tax)	21.3
(11)	Employee Underreported Tax (% of total tax)	5.1

Notes: The table shows measures of underreporting of salaries based on discrepancies between employees' and employers' reports of workers' salaries. Panel A shows the remarkably high prevalence of discrepancies between the two reports. Row (1) shows the percentage of workers who report a salary at least 0.25% smaller than their employers using only individuals who have a single job in the employer statements. Panel B shows how much salary income is underreported. Row (2) sums the discrepancies, row (3) shows the total salary income reported by these individuals' employers, and row (4) shows the total salary income reported by all employers. Row (5) shows the extent of underreporting by evaders by dividing total underreported income (row (2)) by their employers' reported salary (row (3)). Row (6) shows the overall extent of underreporting by dividing underreported income (row (2)) by total reported salary income (row (4)). Rows (7)–(11) repeat this exercise converting the incomes into tax revenues assuming that the worker's salary is his/her taxable income and applying the tax schedule. Since most workers do not have any non-salary income, this approximation will be precise, and since most workers that *do* have non-salary income have positive non-salary income, this approximation will underestimate the effect due to the convexity of the tax schedule.

2.7. Proofs

2.7.1. Proof of Lemma 1 . Using (2.2.2), the distribution of taxable incomes is given by

$$J^*(z) = \begin{cases} \int_0^{z/\left[p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}\right]} \int_0^{z/\left[p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}\right]-\alpha} g(\alpha,\beta)\, d\beta d\alpha & \text{if } z < K_1 \\ \int_0^{z/\left[p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}\right]} \int_0^{z/\left[p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}\right]-\alpha} g(\alpha,\beta)\, d\beta d\alpha & \text{if } K_1 \le z < K_2 \\ \int_0^{z/\left[p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}\right]} \int_0^{z/\left[p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}\right]-\alpha} g(\alpha,\beta)\, d\beta d\alpha & \text{if } K_2 \le z \end{cases}$$

Which features bunching at the kinks. For example, at K_1, $J^*(K_1) = \int_0^{\bar{\delta}_1} \int_0^{\bar{\delta}_1-\alpha} g(\alpha,\beta)\, d\beta d\alpha$, while $\lim_{z\uparrow K_1} J^*(K_1) = \int_0^{\underline{\delta}_1} \int_0^{\underline{\delta}_1-\alpha} g(\alpha,\beta)\, d\beta d\alpha < J^*(K_1)$, and so $B_z(K_1) > 0$. A similar reasoning implies that $B_z(K_2) > 0$.

2.7.2. Proof of Lemma 2 . Using (2.2.2), the salary distribution for workers without non-salary income is given by

$$H^*(s|n^*=0) = \begin{cases} G\left(\frac{s}{p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}}, 0\right) & \text{if } s < K_1 \\ G\left(\frac{s}{p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}}, 0\right) & \text{if } K_1 \le s < K_2 \\ G\left(\frac{s}{p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}}, 0\right) & \text{if } K_2 \le s \end{cases}$$

which features excess bunching at the kinks as $\lim_{s\uparrow K_j} = G\left(\frac{K_j}{p^{1+\varepsilon}(1-\tau_{j-1})^{\varepsilon}}, 0\right) < G\left(\frac{K_j}{p^{1+\varepsilon}(1-\tau_j)^{\varepsilon}}, 0\right)$ for $j = 1, 2$. The excess bunching is given by $B_s(K_j|n^*=0) = G\left(\frac{K_j}{p^{1+\varepsilon}(1-\tau_i)^{\varepsilon}}, 0\right) - \left(\frac{K_j}{p^{1+\varepsilon}(1-\tau_{i-1})^{\varepsilon}}, 0\right)$.

Using (2.2.2) again, the salary distribution for workers with non-salary income is given by

$$H^*(s|n^*>0) = \int_0^s \int_0^\infty g(\alpha(s',\beta),\beta)\, d\beta ds'$$

where

$$\alpha(s,\beta) = \begin{cases} \frac{s}{p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}} & \text{if } s < K_1 \,\&\, \beta \le \underline{\delta}_1 - \frac{s}{p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}} \\ \frac{\beta}{\frac{K_1}{s}-1} & \text{if } s < K_1 \,\&\, \underline{\delta}_1 - \frac{s}{p^{1+\varepsilon}(1-\tau_0)^{\varepsilon}} < \beta \le \bar{\delta}_1 - \frac{s}{p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}} \\ \frac{s}{(1-\tau_1)^{\varepsilon}} & \text{if } s < K_2 \,\&\, \bar{\delta}_1 - \frac{s}{p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}} < \beta \le \underline{\delta}_2 - \frac{s}{p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}} \\ \frac{\beta}{\frac{K_2}{s}-1} & \text{if } s < K_2 \,\&\, \underline{\delta}_2 - \frac{s}{p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}} < \beta \le \bar{\delta}_2 - \frac{s}{p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}} \\ \frac{s}{(1-\tau_2)^{\varepsilon}} & \text{if } \bar{\delta}_2 - \frac{s}{p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}} < \beta \end{cases}$$

Since $\alpha(s,\beta)$ is continuous in both s and β, and since by assumption $g(\alpha,\beta)$ is continuous, $H^*(s|n^*>0)$ is continuous everywhere, including at K_1 and K_2.

2.7.3. Proof of Lemma 3 . From (2.2.2), workers have taxable income K_2 whenever $\underline{\delta}_2 \leq \alpha_i + \beta_i \leq \bar{\delta}_2$. Therefore, for any $s < K_2$, the excess bunching mass of individuals with taxable income K_2 is

$$B^*_{K_2}(s) = \int_{\frac{K_2-s}{p^{1+\varepsilon}(1-\tau_1)^{\varepsilon}}}^{\frac{K_2-s}{p^{1+\varepsilon}(1-\tau_2)^{\varepsilon}}} g\left(\frac{\beta}{\frac{K_2}{s}-1},\beta\right)d\beta \tag{2.7.1}$$

Since the function $\beta/\left(\frac{K_2}{s}-1\right)$ is continuous in β and s, and since $g(\alpha,\beta)$ is smooth by assumption, $B^*_{K_2}(s)$ is continuous in s at all s including K_1.

2.7.4. Proof of Prediction 2 . First note that since real income choices are not distorted by the presence of evasion, and since n^* is increasing in β, individuals' non-salary income in equilibrium will be increasing in β and so we can erform comparative statics with respect to β. Applying the implicit function theorem to (2.2.3),

$$\frac{de^*_0}{d\beta} = \frac{\partial e(0, n-\hat{n}^*_0)}{\partial\beta} - \frac{\partial e(s-\hat{s}, n-\hat{n}^*_s)}{\partial\beta}$$

Since $e_n(0,n-\hat{n}^*_0) = e_n(s-\hat{s}^*, n-\hat{n}^*_s) = \tau$ and $\partial^2 e/\partial(s-\hat{s})\partial(n-\hat{n}) > 0$, it must be the case that $n-\hat{n}^*_s < n-\hat{n}^*_0$. Then, since $\partial e/\partial\beta < 0$ and $\partial^2 e/\partial\beta\partial(n-\hat{n}) > 0$, the first part of the prediction follows. To see the second part, apply the implicit function theorem to the pair of first order conditions $e_s(s-\hat{s}^*, n-\hat{n}^*_s) = e_n(s-\hat{s}^*, n-\hat{n}^*_s) = \tau$ to see that

$$\frac{ds-\hat{s}^*}{d\beta} = \frac{e_{sn}e_{n\beta}}{e_{nn}e_{ss}-e^2_{ns}} > 0$$

where subscripts denote partial derivatives. The inequality follows from the convexity of e and the assumptions that $e_{sn} > 0$ and $e_{n\beta} > 0$.

2.7.5. Proof of Prediction 3. Equation (2.2.2) shows that in the absence of evasion, bunchers at kink $j = 1, 2$ are those for whom$\underline{\delta}_j \leq \alpha_i + \beta_i \leq \bar{\delta}_j$. For the case with evasion, assume for simplicity that $e_0 = 0$, and define

$$\begin{aligned} V(z,\hat{z}) &\equiv \max_{h,q,\hat{s},\hat{n}} U(c,l,q,\hat{s},\hat{n}) \; s.t. \, p(l+q) = z \,\&\, \hat{s}+\hat{n} = \hat{z} \\ &= c - \frac{(\alpha+\beta)^{-1/\varepsilon}}{1+1/\varepsilon} z^{1+1/\varepsilon} - \tilde{e}(z-\hat{z}) \end{aligned}$$

as the maximal utility of earning z and reporting $\hat{z}$, where $\tilde{e}\left(z-\hat{z}\right)=e\left(\frac{\alpha}{\alpha+\beta}\left(z-\hat{z}\right),\frac{\beta}{\alpha+\beta}\left(z-\hat{z}\right)\right)$. Under a linear tax at rate τ, the optimal choices of $z\left(\tau\right)$ and $\hat{z}\left(\tau\right)$ satisfy

$$\begin{aligned} 1-\left(\frac{z\left(\tau\right)}{\alpha+\beta}\right)^{\frac{1}{\varepsilon}}-\tilde{e}_z\left(z\left(\tau\right)-\hat{z}\left(\tau\right)\right) &= 0 \\ -\tau+\tilde{e}_z\left(z\left(\tau\right)-\hat{z}\left(\tau\right)\right) &= 0 \end{aligned}$$

where $\tilde{e}_z=\partial\tilde{e}/\partial\left(z-\hat{z}\right)$ implying that $z\left(\tau\right)=\left(\alpha+\beta\right)\left(1-\tau\right)^{\varepsilon}$ and that $\hat{z}\left(\tau\right)=z\left(\tau\right)-\tilde{e}_z^{-1}\left(\tau\right)$. Individuals who report taxable income at a kink are those for whom the optimal reported income under a linear tax at the lower rate below the kink is above the kink, and the optimal reported income under a linear tax at the higher rate is below the kink:[31]

$$\hat{z}\left(\tau_j\right)\leq K_j\leq\hat{z}\left(\tau_{j-1}\right)\ j=1,2$$

Solving this yields

$$\underline{\delta}_j+\frac{\tilde{e}_z^{-1}\left(\tau_{j-1}\right)}{\left(1-\tau_{j-1}\right)^{\varepsilon}}\leq\alpha_i+\beta_i\leq\bar{\delta}_j+\frac{\tilde{e}_z^{-1}\left(\tau_j\right)}{\left(1-\tau_j\right)^{\varepsilon}}$$

which is a larger range of $\alpha_i+\beta_i$ than in the case without evasion since $\tilde{e}$ is strictly convex. As long as the distribution of $\alpha_i+\beta_i$ is roughly uniform and/or the kink is small, this will mean a larger excess bunching mass at the kink. This continues to be the case when $e_0>0$, though the derivations are slightly more complicated.

2.7.6. Proof of Prediction 5. At salary level s there are $\delta f^*\left(s|n\neq0\right)$ individuals who have $\phi_i=0$ and have chosen salary s as their preferred salary. There are also $\left(1-\delta\right)f^o\left(s\right)$ individuals who have $\phi=\infty$ and happened to receive a salary offer at s. Among the unconstrained individuals, $B_{K_2}^*\left(s\right)$ choose taxable incomes at K_2 as defined in (2.7.1). Among the individuals who are constrained to accept a salary at s despite it not being their preferred salary, $\hat{B}_{K_2}\left(s\right)$ choose taxable incomes at K_2. Since these individuals are constrained in their choices by having a salary away from their preferred salary, fewer of them have taxable incomes that bunch at K_2. Among individuals with $\phi_i=0$, $B_z\left(K\right)$

[31] This argument is made simpler by the absence of income effects under this parameterization of the utility function so that only the marginal tax rate matters. In the presence of income effects the comparison would still be of two linear taxes, but one would have to adjust the intercept of the tax schedule to account for the kink (see Saez, 2010 for further details).

Define the minimum disutility of achieving income z for individuals whose salary choice is unconstrained ($\phi_i = 0$) as

$$\begin{aligned} V^*(z) \quad &\equiv \quad \min_{h,q} \frac{(\alpha+\beta)^{1-\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)}}{1+\frac{1}{\varepsilon}} \left[\alpha\left(\frac{h}{\alpha}\right)^{\sigma} + \beta\left(\frac{q}{\beta}\right)^{\sigma}\right]^{\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)} \\ & \qquad s.t.\, p(h+q) = z \end{aligned}$$

which, solving, yields

$$V^*(z) = \frac{1}{1+\frac{1}{\varepsilon}}\left[\frac{z}{p(\alpha+\beta)}\right]^{\frac{1}{\varepsilon}} \tag{2.7.2}$$

Those who bunch at K_2 are those for whom $(1-\tau_2) \leq \frac{\partial V^*(z)}{\partial z} \leq (1-\tau_1)$, (solving this yields the fourth row of expression (2.2.2)). $\frac{\partial V^*}{\partial z}$ is strictly decreasing and continuous in α and β, so for each α there is an interval of values of β which lead the individual to bunch taxable income at K_2. In the unconstrained case, we have that

$$\frac{\partial^2 V^*(z)}{\partial z \partial \beta} = -\frac{1}{\varepsilon}\frac{1}{\alpha+\beta}\frac{\partial V^*(z)}{\partial z} \tag{2.7.3}$$

For individuals constrained to earn a salary s ($\phi_i = \infty$), the minimum disutility of achieving taxable income $z \geq s$ is

$$\tilde{V}(z) = \frac{(\alpha+\beta)^{1-\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)}}{1+\frac{1}{\varepsilon}}\left[\alpha\left(\frac{s}{\alpha}\right)^{\sigma} + \beta\left(\frac{z-s}{\beta}\right)^{\sigma}\right]^{\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)}$$

and again, those who bunch are those for whom $(1-\tau_2) \leq \frac{\partial \tilde{V}(z)}{\partial z} \leq (1-\tau_1)$, i.e. those for whom

$$1-\tau_2 \leq \left[\frac{\alpha}{\alpha+\beta}\left(\frac{s}{\alpha}\right)^{\sigma} + \frac{\beta}{\alpha+\beta}\left(\frac{z-s}{\beta}\right)^{\sigma}\right]^{\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)-1}\left(\frac{z-s}{\beta}\right)^{\sigma-1} \leq 1-\tau_1$$

And differentiating we get that

$$\frac{\partial^2 \tilde{V}(z)}{\partial z \partial \beta} = -\left\{\left[\frac{1}{\sigma}\left(1+\frac{1}{\varepsilon}\right)-1\right]\left[\frac{1}{\alpha+\beta} + \frac{\sigma-1}{\beta}\frac{\beta\left(\frac{z-s}{\beta}\right)^{\sigma}}{\alpha\left(\frac{s}{\alpha}\right)^{\sigma}+\beta\left(\frac{z-s}{\beta}\right)^{\sigma}}\right] + \frac{\sigma-1}{\beta}\right\}\frac{\partial \tilde{V}(z)}{\partial z} \tag{2.7.4}$$

When $\sigma = 1$, this yields $\frac{\partial^2 \tilde{V}(z)}{\partial z \partial \beta} = -\frac{1}{\alpha+\beta}\frac{1}{\varepsilon}\frac{\partial \tilde{V}(z)}{\partial z} = \frac{\partial^2 V^*(z)}{\partial z \partial \beta}$, while when $\sigma = 1 + 1/\varepsilon$, $\frac{\partial^2 \tilde{V}(z)}{\partial z \partial \beta} = -\frac{1}{\beta}\frac{1}{\varepsilon}\frac{\partial \tilde{V}(z)}{\partial z} < \frac{\partial^2 V^*(z)}{\partial z \partial \beta}$. It is straightforward to show that $\frac{\partial^2 \tilde{V}(z)}{\partial z \partial \beta}$ is strictly decreasing in σ and continuous, so for each α the interval of β that leads to bunching is smaller in the constrained case than in the unconstrained case.

2.7.7. Proof of Prediction 6. In the presence of both information and mismatch effects, the probability of being a taxable income buncher at K_2 conditional on receiving a salary $x \neq K_1$ is

$$p(x) = \bar{\gamma}\frac{\delta f^*(x|n>0)\,B_{K_2}(x|\phi_i=0) + (1-\delta)\,f^o(x)\,B_{K_2}(x|\phi_i=\infty)}{\delta f^*(x|n>0) + (1-\delta)\,f^o(x)} \tag{2.7.5}$$

At K_1, there is firm bunching, so there are an additional $f^0(K_1)\,b_S$ workers with salaries at K_1 compared to x, $B_{K_2}(K_1|\phi_i=\infty)$ of whom have taxable incomes that bunch at K_2, where $f^0(K_1)$ is the counterfactual density at K_1 in the absence of firm bunching. This counterfactual density can be approximated by the observed density at x close to K_1, so that $f^0(K_1) \approx \delta f^*(x|n>0) + (1-\delta)\,f^o(x)$. Combining these, and using the fact that x is close to K_1 so that $B_{K_2}(K_1|\phi_i=\infty) \approx B_{K_2}(x|\phi_i=\infty)$ and $B_{K_2}(K_1|\phi_i=0) \approx B_{K_2}(x|\phi_i=0)$,

$$\begin{aligned} p(K_1) \;=\; & \frac{(\bar{\gamma}+\Delta\gamma)}{(1+b)\left[\delta f^*(x|n>0)+(1-\delta)\,f^o(x)\right]}\Big\{\delta f^*(x|n>0)\,B_{K_2}(x|\phi_i=0) \\ & + (1-\delta)\,f^o B_{K_2}(x|\phi_i=\infty) + b\left[\delta f^*(x|n>0)+(1-\delta)\,f^o(x)\right]B_{K_2}(x|\phi_i=\infty)\Big\} \end{aligned} \tag{2.7.6}$$

Combining (2.7.5) and (2.7.6) yields the result in equation (2.2.7)

2.8. Appendix Figures and Tables

TABLE 10. MERGING THE TAX RETURNS AND THE EMPLOYER STATEMENTS

Result	**2007/08**	**2008/09**	**2009/10**	**2010/11**	**2011/12**
Individual IT Returns	664,425	696,760	681,396	676,699	630,157
Returns with Salary >0	132,209	158,896	164,212	165,897	162,423
NTN & CNIC same on both	51,419	85,081	88,131	83,300	89,029
NTN Match; no CNIC on ES	11,027	1,492	957	665	110
CNIC Match; no NTN on ES	2,404	0	1	0	0
NTN Match; no CNIC on IT	274	13	103	24	11
Total Matched	65,124	86,586	89,192	83,989	89,150
Total Unmatched	67,085	72,310	75,020	81,908	73,273
Match Rate (%)	49.3	54.5	54.3	50.6	54.9

Notes: The table shows the outcome of the procedure used to merge the employer statements (ES) and the income tax returns (IT). Each IT record has a National Tax Number (NTN) identifier, and most also have a Computerised National Identity Card (CNIC) number. Most of the ES records also have at least one of these identifiers, though some have neither. Records are matched whenever the ES record and the IT records contain at least one matching identifier, and no conflicting identifiers. That is, a match occurs whenever a) both the NTN and the CNIC are the same; b) the NTNs are the same but either the ES or the IT record is missing a CNIC; or c) the CNICs are the same but the ES is missing the NTN. A match fails whenever a) the NTNs match, but the IT and ES records have conflicting CNICs; b) the CNICs match but the IT and ES records have conflicting NTNs, or c) when the ES record is missing both the NTN and the CNIC.

TABLE 11. FIRM BUNCHING: ROBUSTNESS

Threshold	TI≠SI	TI≈SI; Unweighted	TI≈SI; PS Reweighted
	1.58	1.02	1.06
0%	(0.142)	(0.143)	(0.161)
	[67,457]	[92,086]	[91,320]
	1.62	1.02	1.32
1%	(0.139)	(0.143)	(0.176)
	[74,197]	[92,373]	[81,672]
	1.68	1.02	1.51
2%	(0.144)	(0.141)	(0.193)
	[75,828]	[92,378]	[80,019]
	1.71	1.02	1.58
3%	(0.145)	(0.140)	(0.196)
	[76,365]	[92,373]	[79,537]
	1.72	1.02	1.73
4%	(0.146)	(0.141)	(0.211)
	[76,601]	[92,345]	[79,958]
	1.72	1.03	1.88
5%	(0.146)	(0.139)	(0.227)
	[76,677]	[92,300]	[80,369]
	1.76	1.00	1.76
$\mathbf{K_{TI} \neq K_{TI}}$	(0.149)	(0.140)	(0.241)
	[76,864]	[91,667]	[78,553]

Notes: The table shows estimates of the normalized excess bunching in the distribution of scaled salaries, b, for different samples of workers, together with the estimate's bootstrapped standard error in round brackets, and the number of observations in the sample in square brackets. Each row uses a different threshold percentage difference between a worker's taxable income and his/her salary to define workers who have "significant" non-salary income or deductions. The first column shows the percentage used as a threshold. The second column shows the excess bunching amongst all workers who file a return. The third column shows the excess bunching amongst workers whose taxable income is approximately the same as their salary, and the final column shows aggregate bunching of salaries amongst workers whose taxable income is significantly different from their salary.

CHAPTER 3

Optimal Income Taxation with Career Effects of Work Effort[1]

ABSTRACT. The literature on optimal income taxation assumes that wage rates are generated exogenously by innate ability and therefore do not respond to behavior and taxation. This is in stark contrast to a large empirical literature documenting a strong effect of current work effort on future wage rates. We extend the canonical Mirrleesian optimal tax framework to incorporate such career effects and provide analytical characterizations that depend on estimable entities. Besides the standard static earnings elasticity with respect to the marginal tax rate, the optimal tax schedule also depends on the elasticity of future wages with respect to current work effort. We explore the empirical magnitude of this "career elasticity" in a meta-analysis of the literature on the returns to work experience and tenure, concluding that a reasonable value for this elasticity lies between 0.2 and 0.4. Calibrating the model to US micro data (under reasonable values of the career elasticity), we present numerical simulations of optimal nonlinear tax schedules that depend on per-period earnings and potentially on age. In the case of age-independent taxation, the presence of career effects make the tax schedule substantially less progressive than in standard models with exogenous wage rates. In the case of age-dependent taxation, career effects create a strong argument for lower taxes on the old, opposite the recommendation in the recent literature on age-dependent taxation. This result reflects both a career incentive effect and an equity effect, where the latter effect arises because increasing earnings over the career path for each ability level imply that, *conditional on earnings*, age and ability are negatively correlated.

[1]We would like to thank Robin Boadway, Peter Diamond, Mike Golosov, Bas Jacobs, Vilen Lipatov, Alan Manning, Emmanuel Saez and Johannes Spinnewijn for discussions and comments. All errors remain our own.

Whatever muscles I have are the product of my own hard work and nothing else

-Evelyn Ashford: Olympic 100m champion

When I was young, I observed that nine out of ten things I did were failures. So I did ten times more work.

-George Bernard Shaw: Nobel laureate in literature

3.1. Introduction

The modern literature on optimal income taxation is cast in the Mirrleesian framework in which innate ability generates a wage rate that is exogenous and therefore unrelated to individual behavior and taxation. This holds both for static versions of the framework (e.g. Mirrlees, 1971; Diamond, 1998; Saez, 2001) and for recent dynamic versions (e.g. Golosov *et al.*, 2007, 2011; Farhi & Werning, 2012) in which the wage rate is allowed to change over time in a potentially non-deterministic fashion, but never depends on behavior. In this literature, earnings in any period of the life cycle respond to taxation only through contemporaneous changes in hours worked. This assumption stands in sharp contrast to a large body of work in labor economics studying the various ways in which current behavior—including work effort—affects future wages. Motivated by this research, we explore the optimal tax implications of breaking the simple mapping between abilities and wages by allowing current hours worked to affect future productivity and wages.

The link between work effort and future wages is widely documented in a vast literature in labor economics. This literature studies the relationship between the wage rate and various measures of work experience, including potential experience (age minus schooling), actual experience, tenure in an individual's current job, and experience lost as a result of job losses (see Blundell & MaCurdy, 1999 and Farber, 1999 for surveys). Conceptually, a variety of mechanisms are likely to be in operation such as improvements in general and firm-specific human capital (Ben-Porath, 1967), improvements in employer-employee matches (e.g. Manning, 2000) and ability signaling effects (Holmström, 1999). In this chapter, we capture all the channels through which current labor supply affects future wage rates in a simple reduced-form relationship, which keeps the otherwise very complicated dynamic optimal taxation problem tractable and allows us to obtain transparent analytical results that depend on empirical entities.

To explore the empirical magnitude of these effects, we conduct a meta-analysis of seventeen empirical studies that permit the derivation of an estimate of the elasticity of future wages with respect to current work effort—the parameter that we show is crucial for optimal income taxation. We find that 80% of the 108 estimates of this "career elasticity" lie between 0.19 and 0.38, implying that an additional 10% of work effort when young raises wages when old by between 2 and 4%. These effects are strong enough to have important qualitative and quantitative implications for optimal tax schedules.

Our chapter also contributes to the recent debate about age-dependent taxation, as reviewed by Banks & Diamond (2011) in the recent *Mirrlees Review*. This work argues that age constitutes a useful tagging device (Akerlof, 1978), which can be used to relax the incentive compatibility constraints of the optimal income tax problem. For instance, applying the static Mirrlees model separately to different age groups, Kremer (2001) argues that earnings distributions and labor supply elasticities are so different across ages that the implied pattern of optimal tax rates would vary greatly by age. More recently, the dynamic optimal tax literature considers this question (Weinzierl, 2011; Golosov *et al.*, 2011; Farhi & Werning, 2012) and finds that age-dependent tax schedules with higher tax rates on older workers are welfare-improving and able to realize most of the gains from a fully optimized history-dependent tax schedule. A key reason for the power of age-dependence in this literature is the fact that the observed wage distribution of older workers features both a higher mean and a higher variance than the wage distribution of younger workers (Weinzierl, 2011). Seen through the lens of the Mirrlees model, this translates to differences in the mean and variance of the ability distribution that creates an equity and insurance argument for higher taxes on the old. What this argument neglects is that the difference in the wage distributions of the young and the old reflects, not differences in exogenous ability, but the fact that the young and the old are observed at different stages of their (endogenous) careers. This is the issue that forms the basis of our chapter, and we show that it can reverse previous conclusions in the literature.[2]

As our framework of analysis, we consider a two-period Mirrlees model in which the wage rate as young equals innate ability while the wage rate as old is a general function of innate ability and hours worked as young. The young and the old have drawn their

[2] Weinzierl (2011) also discusses the importance of modeling the endogeneity of wage paths in order to fully evaluate the case for age-dependent taxation.

abilities from the same underlying ability distribution, but face different wage rates for two reasons. One reason is that effort as young serves as an investment in labor productivity as old (behavioral career effect). The other reason is that, independently of individual behavior, a given innate ability may be associated with an age-varying wage profile rather than a constant wage over the career (what we call a mechanical career effect).[3]

We show that the presence of behavioral career effects provides a plausible microfoundation for the well-documented empirical fact that labor supply elasticities are larger for older workers than for younger workers (e.g. Blundell & MaCurdy, 1999). Since the young are working to raise future wages as well as for consumption in the present while the old are working only to finance consumption in the present, the labor supply of the young is naturally less elastic than the labor supply of the old under the same preferences. Besides these implications for the *own-tax* elasticities of the young and the old, career effects have implications for the *cross-tax* elasticities as, for example, lower taxes on the old induce the young to work harder due to the effort investment effect, what we label the *aspiration effect* in the chapter.

We consider a preference structure allowing us to bypass issues related to savings and capital taxation, and provide analytical characterizations of the optimal taxation of labor earnings that relate in intuitive and transparent ways to existing results without career effects.[4] These characterizations show that the optimal tax schedule can be expressed as a function of long-run earnings elasticities for the young and the old that incorporate the implications of endogenous career paths. Since such long-run earnings responses are not what is captured by the empirical labor supply and taxable income literatures using short-run variation in micro data (as pointed out by, e.g., Piketty & Saez, 2013), we show that the relevant long-run elasticities depend on two underlying sufficient statistics: the standard static earnings elasticity with respect to the marginal net-of-tax rate (as estimated in the enormous taxable income elasticity literature) and the elasticity of future wages with respect to current work effort (the magnitude of which can be inferred from the large literature on the returns to work experience and tenure).

[3]Here we will assume that this wage profile is deterministic, but there is no reason that our framework could not be extended to allow for it to be stochastic.

[4]Consistent with real-world tax policy, we focus on annual tax schedules that involve separate taxation of earnings in different time periods—but may be age-dependent—rather than fully history-dependent tax schedules.

For the case of age-dependent taxation, this framework brings to the fore two important effects that have been ignored in previous optimal tax analyses. First, in the empirically relevant case of increasing wage profiles over the career, an old worker of a given ability level has a higher wage rate and higher earnings than a young worker of the same ability level. As a consequence, an old worker at a given *earnings* level must be of lower ability than young workers at the same earnings level. Therefore, conditional on earnings, age is negatively correlated with ability which creates a classical tagging argument for supplementing an earnings-based income tax with a tax break to older workers. Second, the presence of behavioral career effects create an efficiency argument for lower taxes later in the career, an effect that operates through the own-tax and cross-tax elasticities of labor supply described above. In particular, lower taxes on the old are desirable both because older workers are relatively elastic with respect to their contemporaneous tax rate and because younger workers are elastic with respect to their future tax rate via the aspiration effect. In summary, both the age-ability correlation effect and the elasticity effects call for age-dependent taxation with lower income tax rates on older workers.[5] This is directly opposite to the policy recommendation in the recent optimal tax literature, but is consistent with the policy debate outside economics in which age-dependence is typically discussed in the context of tax breaks for older workers.[6]

When taxes are constrained to be age-independent, we show that the optimal schedule of marginal tax rates can be written as a weighted average of the two optimal age-dependent marginal tax rate schedules. Since earnings increase over the career path, at higher income levels a greater fraction of the population is old and so the weight placed on the old relative to the young in the optimal marginal tax schedule is increasing in income. As the optimal age-dependent marginal tax rates are lower on old workers, the increasing weight on the old makes the optimal age-independent marginal tax rate schedules flatter—less progressive—than in the standard model with exogenous wage paths.

[5]Besides these two effects, a third offsetting effect is driven by the different hazard ratios of the earnings distributions of young and old workers. Optimal marginal tax rates on earnings depend positively on such hazard ratios (see e.g. Saez, 2001, in the context of the standard Mirrlees model), and the empirical fact that earnings distributions of old workers feature higher hazard ratios than earnings distributions of young workers makes it more efficient to tax the old than the young, other things equal. This is precisely the effect that is central to the results in Kremer (2001) and Weinzierl (2011), as discussed above, but in our analysis it is not sufficiently strong to overturn the other arguments calling for lower taxes on the old.

[6]For example, the UK tax system involves limited age-dependence favoring old workers, and the *Mirrlees Review* proposes to go further in this direction.

In order to ascertain the quantitative implications of the new effects we have identified, we carry out numerical simulations based on data for the United States, extending the simulation method set out by Saez (2001) to a setting with career effects. The simulations for age-dependent tax schedules reaffirm the theoretical arguments made above. In a setting with no behavioral career effects (but mechanical career effects generating an increasing wage profile over the life cycle at a given ability), the optimal tax system features a weak degree of age dependence with slightly higher taxes on older workers. However, even very modest behavioral career effects are sufficient to reverse this result and generate lower taxes on older workers. Under realistic assumptions about the strength of career effects (based on our empirical meta-study), it is possible to generate very strong age dependence with much lower taxes on older workers. This result is driven by the age-ability correlation and elasticity effects discussed above.[7] The simulation results for age-independent tax schedules show that even modest career effects can have substantial impacts on optimal marginal tax rates, which are lower and flatter than in the absence of career effects.

The idea that work effort represents an investment in higher future wages (for example via learning by doing) is related to the large literature on human capital investments. Since the implications of standard human capital investments (formal education) for optimal taxation have been explored in earlier work (e.g. Eaton & Rosen, 1980; Bovenberg & Jacobs, 2005), it is important to note that the tax implications of learning by doing are fundamentally different from the implications of education. First, education and work represent two substitutable uses of time, and the key cost of education is therefore the opportunity cost of foregone net-of-tax earnings during education. This implies that education costs are effectively tax deductible in which case income taxation need not distort human capital investments at all (Eaton & Rosen, 1980). By contrast, since learning by doing is a byproduct of work effort, income taxation will always distort this form of human capital investment. Second, formal education is an activity that can be observed and therefore directly subsidized or taxed by the government, whereas learning by doing cannot be separated from labor supply and so cannot receive a separate tax treatment. For both of these reasons, models of optimal taxation with endogenous education are

[7] In particular, these two effects dominate the effect coming from the difference in the earnings distributions of the young and the old (what we will call the hazard ratio effect), which is what drove the previous findings that age dependence should feature higher taxes on older workers.

conceptually very different from our framework and do not shed light on the issues that we highlight in this chapter. As far as we are aware, the only previous chapter that allows for learning-by-doing effects in the context of optimal income taxation is Krause (2009), who focuses on the implications of such effects for the no-distortion-at-the-top result in the context of a two-type Stiglitz (1982) model.

We will proceed as follows. Section 3.2 presents the setting and shows the implications of career effects for earnings elasticities. Section 3.3 characterizes optimal income tax schedules and discusses the implications of career effects for both age-dependent and age-independent taxation. Section 3.4 investigates empirically the career effect of work effort based on a meta-analysis of the literature on experience and tenure effects. Section 3.5 presents numerical simulations that demonstrate the quantitative importance of career effects for optimal tax policy, and finally section 3.6 concludes.

3.2. The Setting

3.2.1. Individuals. We analyze the simplest possible setting that allows us to explore the implications of career effects for optimal tax schedules. Individuals live for 2 periods, $i \in \{y, o\}$, work in both of them and at any point in time there is a continuum of mass 1 of individuals of each age alive. They have time separable preferences with no discounting and their per-period utility is quasi-linear and given by $u(c_i, l_i) = c_i - \frac{1}{1+1/e} l_i^{1+1/e}$. This formulation has the virtue that individuals will not save and so we can focus on the analysis of wage effects without the additional complication of saving effects.

In the first period of life, individuals are paid according to their innate ability n, the distribution of which is given by the cdf $F(n)$. Therefore, earnings when young are $z_y = n l_y(n)$. Our key innovation is to allow the second-period wage to depend both on innate ability and on the first period's effort choice. We allow this effect to manifest itself in a very general way, merely positing that the wage rate when old ω is a general function of innate ability and first-period effort, i.e. $\omega = \omega(n, l_y)$. Earnings when old are then given by $z_o = \omega(n, l_y) l_o$. The responsiveness of the wage rate when old to innate ability may be captured by the elasticity $\eta = \frac{\partial \omega}{\partial n} \frac{n}{\omega}$ and reflects the *mechanical career effect* of higher ability on the life-cycle profile of wages. The responsiveness of wages when old to effort when young is captured by the elasticity $\delta = \frac{\partial \omega}{\partial l_y} \frac{l_y}{\omega} = \frac{\partial \omega}{\partial z_y} \frac{z_y}{\omega}$ and reflects the *behavioral career effect* due to the investment component of work effort as young. When we turn to simulations of the optimal tax schedules in section 3.5, we will assume that we are in the empirically plausible case where $\eta, \delta \geq 0$, however note that this restriction is not necessary for our derivations of the optimal tax schedules.

Since there are no savings, consumption at age i is simply equal to earnings net of income taxes at that age, i.e. $c_i = z_i - T_i(z_i)$. The income tax liability at age i, $T_i(z_i)$, depends on earnings at that age (but not on earnings at other ages) and possibly on age itself (as the $T_i(.)$ function is allowed to vary with i). This is consistent with real-world tax schedules, which are always based on annual income and sometimes feature aspects of age-dependence (see, for example, the Mirrlees Review for a description of age-dependence in the UK tax system).

Lifetime utility is given by

(3.2.1)
$$U(z_y, z_o) = z_y - T_y(z_y) - \frac{1}{1+1/e}\left(\frac{z_y}{n}\right)^{1+1/e} + z_o - T_o(z_o) - \frac{1}{1+1/e}\left(\frac{z_o}{\omega(n, z_y/n)}\right)^{1+1/e}$$

which has first-order conditions for earnings chosen when young and when old given by

(3.2.2)
$$1 - \tau_y(z_y) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}} \frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \frac{\delta}{z_y} = 0$$

and

(3.2.3)
$$1 - \tau_o(z_o) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}} \frac{1}{\omega} = 0$$

where $\tau_i(z) \equiv T_i'(z)$ is the marginal tax rate on earnings in period i.

3.2.2. Earnings Elasticities. To facilitate interpretation of our main results, this section starts by characterizing the relationship between the strength of career effects and earnings elasticities for the young and the old.[8] At the extreme, when there are no behavioral career effects ($\delta = 0$), this model reduces to a simple two-period version of a standard optimal income tax model like that studied in Diamond (1998). In particular, the young are responsive only to the tax schedule they face when young even though they know

[8]Throughout the chapter, we focus on earnings elasticities (including hours-worked and wage-rate effects) rather than hours-worked elasticities, because it is the former elasticity concept that matters directly for optimal tax schedules. However, the main qualitative properties of earnings elasticities that we characterize in this section also applies to hours-worked elasticities.

the tax schedule they will face when old, and similarly for the old. This is because their behavior when young does not affect the decision-making problem as old, and vice versa. Moreover, it is easy to see from the first-order conditions (3.2.2) and (3.2.3) that the elasticity of earnings at age i with respect to the marginal net-of-tax rate at that age, $1-\tau_i$, is given by the utility parameter e for both age groups. However, when we introduce career effects through $\delta > 0$, this changes.

We define the elasticity of earnings at age i with respect to the marginal net-of-tax rate at age j as $E_{ij} \equiv \frac{dz_i}{d(1-\tau_j)} \frac{1-\tau_j}{z_i}$. Applying the implicit function theorem to the pair of first-order conditions (3.2.2) and (3.2.3), Appendix 3.7 shows that the earnings elasticities can be expressed as

$$\begin{pmatrix} E_{yy} & E_{yo} \\ E_{oy} & E_{oo} \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} e & e\,(1+e)\,\delta \frac{z_o(1-\tau_o)}{z_y(1-\tau_y)} \\ e\,(1+e)\,\delta & e\left[1+\delta\,(1+\delta)\,(1+e)\,\frac{z_o(1-\tau_o)}{z_y(1-\tau_y)}\right] \end{pmatrix} \tag{3.2.4}$$

where $\kappa \equiv 1+\delta\,(1-e\delta)\,(1+e)\,\frac{z_o(1-\tau_o)}{z_y(1-\tau_y)}$. The elasticities E_{yy} and E_{oo} are *contemporaneous* earnings elasticities of the young and the old with respect to the marginal net-of-tax rates faced at those respective ages, while E_{yo} and E_{oy} are *intertemporal* earnings elasticities of the young and the old that reflect the presence of career effects. The elasticity E_{yo} reflects what we refer to as the *aspiration effect*: since part of the return to current work effort is higher future wages, and individuals anticipate the rate at which those future wages will be taxed, a higher tax rate later in life reduces the career investments made through work effort earlier in life. The elasticity E_{oy} reflects what we refer to as the *accumulation effect*: a higher tax rate on the young reduces work effort and therefore earnings by the young, which has a negative knock-on effect on the wage rate and labor supply of those individuals when they become old.

In the following, we present three lemmas that clarify the precise link between the career effect δ and the size of earnings elasticities. The proofs of these lemmas are provided in appendix 3.7. The first lemma shows how the contemporaneous responsiveness of the two age groups is affected by the presence of career effects:

LEMMA 6. *In the absence of behavioral career effects, $\delta = 0$, the contemporaneous earnings elasticities of the young and the old are given by $E_{yy} = E_{oo} = e$. In the presence of behavioral* *career effects, $\delta > 0$, the contemporaneous earnings elasticity of the young is lower while that of the old is larger than in the absence of such effects, i.e.* $E_{yy} < e$ and $E_{oo} > e$.

Intuitively, the young are working both for current wages (taxed at rate τ_y) and to raise their wages when old (taxed at rate τ_o), and so their earnings are naturally less elastic to their tax rate as young than is implied by the standard static elasticity e. Meanwhile, the earnings of the old respond to the tax rate when old both through a standard static hours-of-work response governed by the e-parameter and through a dynamic wage-rate response coming from the effect of the tax rate when old on the incentive while young to invest in higher wages as old. Notice that these earnings elasticities (and those discussed below) reflect full dynamic effects on earnings at different ages by taxpayers who plan their entire life cycle profile of earnings, perfectly anticipating the tax schedule faced in each period. These are, of course, the relevant elasticities to consider for the optimal tax analysis that follow, which focuses on the optimal tax policy by a government that can fully commit to future tax rates.

Next, we turn to the implications of career effects for the aspiration and accumulation elasticities:

LEMMA 7. *In the absence of behavioral career effects, $\delta = 0$, the aspiration and accumulation elasticities are zero, i.e. $E_{yo} = E_{oy} = 0$. In the presence of behavioral career effects, $\delta > 0$, the aspiration and accumulation elasticities are positive and always increasing in the strength of the career effect, i.e. $\frac{\partial E_{yo}}{\partial \delta} > 0$ and $\frac{\partial E_{oy}}{\partial \delta} > 0$.*

The intuition behind these results follows naturally from the fact that, in this model, it is precisely the effect of current work effort on future wage rates that creates an intertemporal link between taxation and earnings across different periods. With positive career effects of work effort, earnings in one period respond positively to the net-of-tax rate in another period, and the size of this response is increasing in the size of the career effect δ.

The elasticities considered so far measure earnings responses as young or old to the tax rate in one period of life taking as given the tax rate in the other period. It is useful to also consider *total* earnings responses by the young and the old to a change in the tax rate in both periods of life. Defining the total elasticity of earnings at age i as $E_i \equiv E_{iy} + E_{io}$, we can state the following:

LEMMA 8. *In the presence of behavioral career effects, $\delta > 0$, the total elasticity of earnings in each period is larger than the standard static elasticity, i.e.*

$$E_y \equiv E_{yy} + E_{yo} \quad > \quad e$$

$$E_o \equiv E_{oy} + E_{oo} \quad > \quad e$$

Moreover, with $\delta > 0$, the total elasticity of earnings as old is larger than the total elasticity of earnings as young, i.e.

$$E_o > E_y$$

These results demonstrate that career effects increase the overall responsiveness of earnings to income taxation and therefore exacerbate the efficiency costs of taxation. Moreover, the degree to which career effects increase the responsiveness of earnings is stronger for the old than for the young. This last result not only provides an interesting and plausible micro-foundation for the often reported finding that labor supply and earnings elasticities are larger for the old than for the young (see, for example Blundell & MaCurdy, 1999),[9] it also has potentially important implications for optimal tax structure and in particular the desirability and design of age-dependent taxes.

3.2.3. The Government. We consider a government imposing an "annual" income tax that may or may not depend on the age of the taxpayer. That is, an individual's tax liability in a given period depends exclusively on within-period income and possibly on age. This is analytically and conceptually different from considering a government choosing fully history-dependent tax schedules in which an individual's tax liability when old may depend directly on income earned when young. We focus on annual age-dependent tax schedules rather than fully history-dependent schedules, because the former is empirically more relevant: real-world income tax systems are based on annual time-separable tax liability and occasionally involves some age-dependence, but are in general not history dependent.[10] While at present, age-dependence in the income tax system is used either in a very limited fashion in some countries (e.g. United Kingdom) or not at all in other

[9]The relationships in Lemma 8, which are stated in terms of earnings elasticities, also apply to hours-of-work elasticities.

[10]There is some history dependence in social security systems, which matters for retirement decisions. But here we focus on income taxation and do not model retirement.

countries (e.g. United States), it is interesting to analyze because of several recent proposals to introduce age as a tagging device in tax systems. We characterize optimal tax policy both when full age dependence is allowed (general schedules $T_y(z), T_o(z)$) and when no age dependence is allowed (schedules $T_y(z) = T_o(z) = T(z)\ \forall z$). We assume throughout that the government can fully commit to future tax rates.

In the case of age-dependence, the government chooses tax schedules for the young and the old $T_y(z), T_o(z)$ to maximize social welfare subject to incentive compatibility constraints and a revenue-raising constraint, i.e.

$$\max_{T_y(z),T_o(z)} \int_0^\infty \Psi\left[U\left(z_y(n), z_o(n)\right)\right] dF(n)$$

$$\begin{aligned} s.t.\quad \{z_y(n), z_o(n)\} &\in \arg\max U(z_y, z_o)\ \forall n \\ \int_0^\infty T_y(z_y(n))\, dF(n) + \int_0^\infty T_o(z_o(n))\, dF(n) &\geq R \end{aligned}$$

where $\Psi[\cdot]$ is an additively separable social welfare function defined over the lifetime utility of individuals, R is an exogenous revenue requirement, and the size of each generation has been normalized to 1. The government's redistributive tastes may be captured by social welfare weights equal to the social marginal utility of income for different individuals expressed in terms of the marginal value of public funds. For an individual of ability n, the social welfare weight is defined as $g(n) \equiv \Psi'\left[U\left(z_y(n), z_o(n)\right)\right]/\lambda$ where λ is the Lagrange multiplier on the government budget constraint, the marginal value of public funds. It will be useful to translate this welfare weight from being a function of ability to being a function of income, so we also define $g_y(z) \equiv \Psi'\left[U\left(z, z_o(z)\right)\right]/\lambda$ and $g_o(z) \equiv \Psi'\left[U\left(z_y(z), z\right)\right]/\lambda$, where $z_o(z)$ are the equilibrium earnings when old of an individual who earns z when young and $z_y(z)$ are the equilibrium earnings when young of an individual who earns z when old. $z_y(z), z_o(z)$ are increasing functions of z as long as $z_y(n), z_o(n)$ are increasing functions of n. As in the standard Mirrlees model, the condition that $z_y(n), z_o(n)$ are increasing in n is necessary and sufficient to ensure that a given path for $z_y(n), z_o(n)$ can be implemented by a truthful mechanism or, equivalently, by a nonlinear tax system. The analytical characterization in section 3.3 assumes that this condition is satisfied while section 3.5 verifies this numerically.

3.3. Optimal Tax Schedules

This section characterizes analytically the implications of career effects for the optimal non-linear tax schedule in the age-dependent and age-independent cases. We derive optimal tax formulas using both Hamiltonian and tax perturbation approaches, where the latter is particularly useful for facilitating economic intuition about the role of different effects. The optimal marginal tax rates are expressed in terms of entities that are observable or estimable in the manner of Diamond (1998) and Saez (2001), which lends itself naturally to a calibration exercise as considered in section 3.5. As we describe in detail below, the implications of behavioral career effects for optimal income taxation can be split into *elasticity effects* coming from how careers affect the responsiveness of earnings by the young and the old to taxes, a *welfare weight effect* coming from how careers affect the social marginal utilities of income of the young and the old, and a *hazard ratio effect* coming from how careers generate different earnings distributions for the young and the old.

3.3.1. Optimal Age-Dependent Taxes.

In this section we characterize the optimal age-dependent, nonlinear income tax schedule $\{T_y(z), T_o(z)\}$ with corresponding marginal tax rate schedules $\{\tau_y(z), \tau_o(z)\}$. We can show:

PROPOSITION 4. *The optimal age-dependent tax schedule, $T_i(z)$ at age $i \in \{y, o\}$, is associated with marginal tax rates*

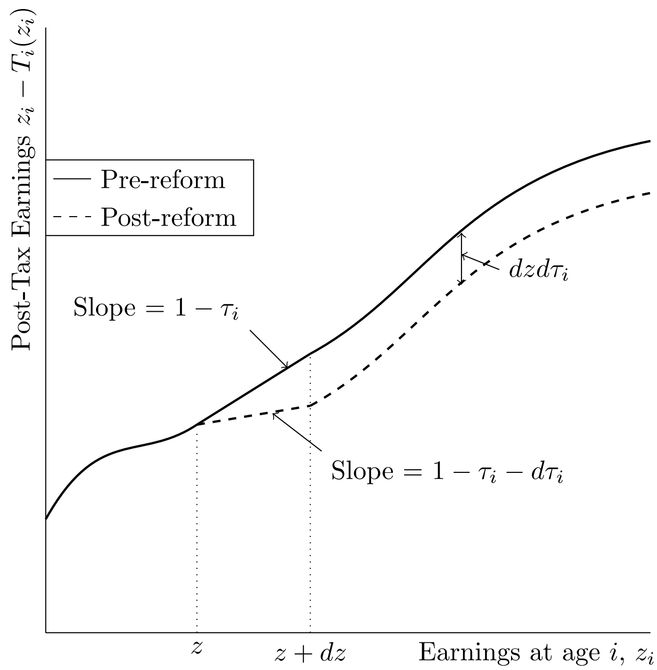
$$\frac{\tau_i(z)}{1-\tau_i(z)} = A_i(z)\, B_i(z)\, C_i(z) \tag{3.3.1}$$

where (for $i \neq j$) we have

$$\begin{aligned} A_i(z) &= \left\{ E_{ii} + E_{ji} \frac{\tau_j(z_j(z))\, z_j(z)}{\tau_i(z)\, z} \right\}^{-1} \\ B_i(z) &= \frac{\int_z^\infty [1 - g_i(z')]\, dH_i(z')}{1 - H_i(z)} \\ C_i(z) &= \frac{1 - H_i(z)}{z h_i(z)} \end{aligned}$$

at any earnings level z. In these expressions, $H_i(.)$ and $h_i(.)$ denote the equilibrium cdf and pdf, respectively, of earnings at age i.

FIGURE 29. NON LINEAR TAX RATE PERTURBATION



Proof. Here we prove the result directly using a tax perturbation method (as first developed by Piketty, 1997; Saez, 2001), first for the young and then for the old as this illustrates the intuition for the results better. A technically more rigorous proof based on the Hamiltonian approach is found in appendix 3.8 alongside a proof that the two methods produce equivalent results in the context of our model. For the young and the old separately, consider a small perturbation around the optimal tax schedule as depicted in Figure 1. The perturbation increases the marginal tax rate by a small amount $d\tau_i$ at age i on incomes falling in a small band $(z, z + dz)$ but is otherwise left unchanged.

The tax schedule of the young. We first consider the perturbation in the tax schedule of the young. The marginal tax rate increase $d\tau_y$ in the small band $(z, z + dz)$ has a mechanical effect on tax revenue and welfare for all young individuals above z as well as two behavioral effects on those with earnings between z and $z + dz$ as young. We proceed to analyze the three effects separately:

Mechanical Welfare Effect. All young taxpayers with earnings above z pay $d\tau_y dz$ more in taxes (holding behavior constant), which creates a mechanical revenue gain for the government but reduces the utility of those individuals. The net social welfare effect of the mechanical tax increase of a young individual with income z' is given by $d\tau_y dz \cdot [1 - g_y(z')]$. Hence, the total mechanical effect on social welfare is given by

$$\Delta_y^M = d\tau_y dz \cdot \int_z^\infty [1 - g_y(z')] \, dH_y(z')$$

Contemporaneous Earnings Effect. Using the definition of the contemporaneous earnings elasticity of the young E_{yy} in section 3.2.2, each young person in the band $(z, z + dz)$ reduces earnings by $-E_{yy} \cdot \frac{d\tau_y}{1-\tau_y(z)} \cdot z$. Multiplying the earnings response by the marginal tax rate $\tau_y(z)$, we get the change in tax liability by each individual in this band. As there are $h_y(z)\,dz$ young individuals in the band, the total effect of contemporaneous earnings responses on tax revenue is given by

$$\Delta_y^E = -d\tau_y dz \cdot z h_y(z) \cdot E_{yy} \cdot \frac{\tau_y(z)}{1 - \tau_y(z)}$$

Accumulation Effect. The labor supply response of young workers located in the band $(z, z + dz)$ affects human capital accumulation and therefore the wage rate and earnings of those young workers when they become old. As established earlier, a given tax system is associated with a mapping between earnings as young and earnings as old, so that a person with earnings z as young has earnings $z_o(z)$ as old. This implies that changing the tax rate on young workers at income level z has an accumulation effect on old workers at income level $z_o(z)$. Using the definition of the accumulation elasticity E_{oy}, an old person at $z_o(z)$ reduces earnings by $-E_{oy} \cdot \frac{d\tau_y}{1-\tau_y(z)} \cdot z_o(z)$. The number of old workers whose earnings change as a result of this accumulation effect (those in the band $(z_o(z), z_o(z + dz))$ of the distribution $h_o(z_o)$) is equal to the number of young workers who changed their labor supply in response to the higher tax rate on the young (those in the band $(z, z + dz)$ of the distribution $h_y(z_y)$), i.e. we have $h_o(z_o(z)) \frac{dz_o}{dz} dz = h_y(z)\,dz$, and therefore the total effect on tax revenue due to the accumulation effect on all old workers affected is given by

$$\Delta^{AC} = -d\tau_y dz \cdot z_o(z)\, h_y(z) \cdot E_{oy} \cdot \frac{\tau_o(z_o(z))}{1 - \tau_y(z)}$$

Optimality. At the optimal tax schedule, there should be no first-order welfare effect of this perturbation, and so we have

$$\Delta_y^M + \Delta_y^E + \Delta^{AC} = 0$$

Inserting the above expressions and rewriting gives the following optimality condition on the tax schedule for the young

$$\frac{\tau_y(z)}{1-\tau_y(z)} = \left\{E_{yy} + E_{oy} \cdot \frac{\tau_o(z_o(z))\, z_o(z)}{\tau_y(z)\, z}\right\}^{-1} \cdot \frac{\int_z^\infty \left[1 - g_y(z')\right] dH_y(z')}{z h_y(z)}$$

which, after multiplying and dividing by $1 - H_y(z)$, is equivalent to the expression in Proposition 4 for $i = y$.

The tax schedule of the old. As in the tax perturbation for the young, the marginal tax rate increase on the old $d\tau_o$ in the band $(z, z+dz)$ gives rise to a mechanical welfare effect above z along with two behavioral effects on those between z and $z + dz$ as old. The mechanical welfare effect on the old is analogous to the expression for the young:

$$\Delta_o^M = d\tau_o dz \cdot \int_z^\infty \left[1 - g_o(z')\right] dH_o(z')$$

There is also a contemporaneous earnings effect on the old taking the same form as for the young:

$$\Delta_o^E = -d\tau_o dz \cdot z h_o(z) \cdot E_{oo} \cdot \frac{\tau_o(z)}{1-\tau_o(z)}$$

Finally, instead of the accumulation effect of the tax perturbation for the young, we have an aspiration effect of the tax perturbation for the old.

Aspiration Effect. The higher tax rate on old workers in the earnings band $(z, z + dz)$ discourages young workers who anticipate being in this band when old from investing in future productivity and earnings. Using the mapping between earnings as young and earnings as old, this behavioral effect on the young occurs in the earnings band $(z_y(z), z_y(z + dz))$. The change in earnings by each young worker who is affected equals $-E_{yo} \cdot \frac{d\tau_o}{1-\tau_o(z)} \cdot z_y(z)$. The number of young workers affected (those in the band $(z_y(z), z_y(z + dz))$ of the distribution $h_y(z_y)$) is equal to the number of old workers facing a higher marginal tax rate (those in the band $(z, z + dz)$ of the distribution $h_o(z_o)$), so that $h_y(z_y(z)) \frac{dz_y}{dz} dz = h_o(z)\, dz$. This implies that the total effect on tax revenue due to the aspiration effect can be written as

$$\Delta^{AS} = -d\tau_o dz \cdot z_y(z)\, h_o(z) \cdot E_{yo} \frac{\tau_y(z_y(z))}{1-\tau_o(z)}$$

Optimality. At the social optimal, we have

$$\Delta_o^M + \Delta_o^E + \Delta^{AS} = 0$$

which gives the expression in Proposition 4 for $i = o$. $\square$

We have thus characterized the optimal tax schedule in terms of two expressions that share several qualitative features with the standard formulas in Diamond (1998) and Saez (2001), but with some important differences that bear fleshing out. We will discuss these in the context of their implications for the optimal form and degree of age dependence in the tax system.

3.3.2. Age Dependence in the Optimal Tax System. The existence of career effects of work effort has implications for all three terms in the optimal income tax formula (3.3.1): the inverse elasticity term A_i, the welfare weight term B_i and the hazard ratio term C_i. Considering each of these terms separately, we now discuss the implications of career effects for optimal income tax structure. We emphasize how career effects change the three key terms in different ways for the young and the old, and therefore have important effects on the optimal form and degree of age dependence in the tax system.

The *elasticity effect* of careers operates through $A_y(z)$ and $A_o(z)$. For the taxation of old workers ($A_o(z)$ term), Lemmas 6 & 7 show that career effects $\delta > 0$ give rise to a contemporaneous earnings elasticity for the old that is larger than the standard static elasticity, $E_{oo} > e$, as well as a positive aspiration elasticity for the young, $E_{yo} > 0$. The combination of these effects imply $A_o(z) < 1/e$, so that the inverse elasticity term for old is always smaller than in standard models without career effects. This calls for lower taxes on the old, other things equal. For the taxation of young workers ($A_y(z)$ term), Lemmas 6 & 7 show that $\delta > 0$ implies a contemporaneous earnings elasticity for the young that is smaller than the standard elasticity, $E_{yy} < e$, along with a positive accumulation elasticity on the old, $E_{oy} > 0$. Hence, depending on the magnitudes of these elasticities, $A_y(z)$ may be either below or above $1/e$. Due to fact that the elasticities E_{yy}, E_{oy} (see equation 3.2.4) and the weighting term on E_{oy} in the optimal tax formula are endogenous to the tax system itself, it is not possible to analytically determine if $A_y(z)$ is smaller or greater than $1/e$. Nevertheless, our numerical simulations (discussed in section 3.5) show that $A_y(z) \gtrsim 1/e$ under a wide range of reasonable parameter assumptions, so that the elasticity effect of careers calls for either unchanged or higher taxes on the young, other things equal. The combination of these insights imply that the elasticity effect on its own calls for age-dependent taxes with lower taxes on the old than on the young, conditional on earnings.

The *welfare weight effect* of careers operates through the terms $B_y(z)$ and $B_o(z)$. In the discussion, it is useful to denote by $G_i(z)$ the average social welfare weight on individuals of age i with earnings above z, so that we may write $B_i(z) = 1 - G_i(z)$. When considering the effect of age on the average social welfare weight $G_i(z)$, notice first that the social welfare weight on any given individual is a function of her lifetime utility which depends on her innate ability, but not on her age. However, the average social welfare weight over the earnings segment (z, ∞) is not independent of age, because this earnings segment is associated with different ability segments for the young and the old due to career effects. Since earnings profiles are increasing over the life cycle, the pool of old workers with earnings above z consists of all those whose earnings were above z as young *and also* some individuals whose earnings were below z as young. Given that earnings are increasing in ability n conditional on age (the condition for implementability of the direct mechanism), those below z as young must be of lower ability than those above z as young. Denoting the average welfare weight among workers who are below z as young but above z as old by $G_y(z^-)$, it follows that $G_y(z^-) > G_y(z)$ under concave social preferences. The average social welfare weight on old workers above z can then be written as $G_o(z) = s \cdot G_y(z) + (1-s) \cdot G_y(z^-) > G_y(z)$ for $s \in (0,1)$. Intuitively, with increasing earnings profiles over the career path at each ability n, older workers in a given earnings range are, on average, of lower ability than young workers in the same earnings range (age and ability are negatively correlated, conditional upon earnings), and therefore the average social welfare weight on the old is larger than on the young. This effect implies $B_o(z) < B_y(z)$, and so the welfare weight effect, like the elasticity effect discussed above, calls for age-dependent taxes with lower taxes on the old than on the young, conditional on earnings.

Finally, the *hazard ratio effect* of careers operates through the terms $C_y(z)$ and $C_o(z)$. These hazard ratios can be seen as measures of the thickness of the earnings distribution above a cutoff z for the young and the old, respectively. As an example, if earnings are distributed according to the Pareto distribution, these ratios are equal to the inverse of the Pareto parameter and measure the thickness of the upper tail. In our model, the presence of increasing earnings profiles over the career create an earnings distribution for older workers with a thicker upper tail than for younger workers, which implies $C_o(z) > C_y(z)$ at least for a high enough z. This prediction is borne out by the data (hazard ratios under actual tax systems) and by our numerical simulations below (hazard ratios under the optimal tax system), in which the hazard ratio is larger for older than for younger workers, except at very low levels of earnings. On its own, this effect calls for higher taxes on the old than on the young, conditional on earnings, and therefore works to offset the elasticity and welfare weight effects described above. This hazard ratio effect is what drives the strong age-dependence results in Weinzierl (2011). In our framework, it is not possible to establish analytically whether the hazard ratio effect (calling for higher taxes on the old) is able to dominate the elasticity and welfare weight effects (calling for lower taxes on the old), and so we turn to numerical simulations based on U.S. micro data to explore this in section 3.5.

3.3.3. Optimal Age-Dependent Top Tax Rates. Assuming that the upper tails of the earnings distributions for the young and the old are both Pareto distributed (with potentially different Pareto parameters), the optimal top marginal tax rates depend on career effects in a particularly simple way. We state the following proposition

PROPOSITION 5. *Suppose that for very high incomes, the earnings of the young and the old are distributed according to Pareto distributions with Pareto parameters a_y and a_o respectively. Suppose further that the welfare weights on the young and the old converge to $\bar{g}_y$ and $\bar{g}_o$ and that the elasticities E_{ij}, $i, j \in \{y, o\}$ converge to constant values denoted by $\bar{E}_{ij}$. Then the optimal top marginal tax rates $\bar{\tau}_i$ on the young ($i = y$) and the old ($i = o$) are given by*

$$\frac{\bar{\tau}_i}{1-\bar{\tau}_i} = \frac{1-\bar{g}_i}{a_i\left[\bar{E}_{ii} + \bar{E}_{ji}\frac{a_j/(a_j-1)}{a_i/(a_i-1)}\frac{\bar{\tau}_j}{\bar{\tau}_i}\right]} \tag{3.3.2}$$

where $i, j \in \{y, o\}$, $i \neq j$.

PROOF. To prove the proposition, we show that each of the components of equation (3.3.1) converges to a constant. $B_i(z)$ and $C_i(z)$ are straightforward. Clearly, if the welfare weights converge to $\bar{g}_i$ then $B_i(z) \to 1-\bar{g}_i$. It is a property of the Pareto distribution that $\left[1 - H_i(z)\right] / \left[z h_i(z)\right] = 1/a_i$ so $C_i(z) \to 1/a_i$. To establish the limiting value of $A_i(z)$ we use the property of the Pareto distribution with Pareto parameter a_i that $\mathsf{E}\left[z|z > x\right] = \frac{a_i}{a_i - 1}x$. For individuals in age group i the limiting value of the ratio of their earnings when in the other age group j to their current earnings is $\lim_{z\to\infty} \frac{z_j(z)}{z} = \lim_{x\to\infty} \mathsf{E}\left[z_j|z_j > x\right] / \mathsf{E}\left[z_i|z_i > x\right] = \frac{a_j/(a_j-1)}{a_i/(a_i-1)}$. Combined with the assumption that the elasticities E_{ij} converge to constant values this implies that $A_i(z) \to \bar{A}_i \equiv \left[\bar{E}_{ii} + \bar{E}_{ji} \frac{a_j/(a_j-1)}{a_i/(a_i-1)} \frac{\bar{\tau}_j}{\bar{\tau}_i}\right]^{-1}$. Combining these pieces establishes the result in equation (3.3.2). □

Equation (3.3.2) highlights the three conceptual effects discussed in section 3.3.2 in a very simple way. The *welfare weight effect* is captured by the term $1 - \bar{g}_i$ (where we have $\bar{g}_o > \bar{g}_y$ since increasing career-earnings profiles imply that, conditional on earnings, the old have lower abilities than the young), the *hazard ratio effect* is captured by the inverse of the Pareto parameter $1/a_i$ (where we have $a_y > a_o$ since increasing career-earnings profiles create a thicker upper tail in the earnings distribution of the old than in the earnings distribution of the young), and finally the *elasticity effect* is captured by the bracketed term in the denominator (where career effects imply $\bar{E}_{oo} > \bar{E}_{yy}$ and $\bar{E}_{yo} > \bar{E}_{oy}$, favoring lower taxes on the old). Note also that, in the limit where $z \to \infty$, the welfare weights on both age groups will asymptote to zero under standard concave social welfare functions, and so the welfare weight effect would not support any age-dependence at the limit. Therefore, at very high levels of earnings, optimal age dependence reflects a simple trade-off between the relative Pareto parameters—the key mechanism in previous work arguing for higher marginal rates on the old (Kremer, 2001; Weinzierl, 2011)—and career incentive effects which tend to call for lower marginal tax rates on the old as discussed above.

3.3.4. Age-Independent Taxes. As current tax systems in the world tend to make limited or no use of explicit age-dependence, it is of obvious interest to consider whether the career effects we introduce have any bite in influencing optimal age-independent tax schedules. This section therefore characterizes the optimal age-independent, nonlinear income tax schedule $T(z)$ with corresponding marginal tax rate schedule $\tau(z)$. We will see that it is still possible to express the optimal tax formula in terms of observable quantities and elasticities, and that the key effects discussed above are still present and affect the level and profile of marginal tax rates. In this setting, we have

PROPOSITION 6. *The optimal age-independent tax schedule $T(z)$ is associated with marginal tax rates*

$$\frac{\tau(z)}{1-\tau(z)} = \frac{\alpha(z)\,B_y(z)\,C_y(z)+[1-\alpha(z)]\,B_o(z)\,C_o(z)}{\alpha(z)\,A_y(z)^{-1}+[1-\alpha(z)]\,A_o(z)^{-1}} \tag{3.3.3}$$

where $\alpha(z)\equiv h_y(z)\,/\,[h_y(z)+h_o(z)]$ is the proportion of individuals with income z who are young and $A_y(z)$, $A_o(z)$, $B_y(z)$, $B_o(z)$, $C_y(z)$ and $C_o(z)$ are as defined in Proposition 4.

Proof: Again, we prove the result directly using the perturbation method, leaving the Hamiltonian method and the demonstration of their equivalence for appendix 3.8. The perturbation that we consider is similar to the one depicted in Figure 1, except that it pertains to the unique tax schedule faced by both the young and the old. Hence, the marginal tax rate on both the young and the old is increased by a small amount $d\tau$ in a small earnings band $(z, z+dz)$. We now characterize the social welfare effects of this tax reform.

Mechanical Welfare Effect. All taxpayers with earnings above z face a mechanical increase in tax liability of $d\tau dz$. For a young individual with earnings $z'>z$ the social value of this is given by $d\tau dz\cdot[1-g_y(z')]$, while for an old individual at $z'>z$ the social value of this equals $d\tau dz\cdot[1-g_o(z')]$. The total mechanical welfare effect is therefore given by

$$\Delta^M = d\tau dz\cdot\left\{\int_z^\infty[1-g_y(z')]\,dH_y(z')+\int_z^\infty[1-g_o(z')]\,dH_o(z')\right\}$$

Contemporaneous Earnings Effects. In the band $(z, z+dz)$, each young person reduces earnings by $-E_{yy}\cdot\frac{d\tau}{1-\tau(z)}\cdot z$ while each old person reduces earnings by $-E_{oo}\cdot\frac{d\tau}{1-\tau(z)}\cdot z$. The total tax revenue implications of these earnings response equal

$$\Delta^E = -d\tau dz\cdot z\cdot\{h_y(z)\cdot E_{yy}+h_o(z)\cdot E_{oo}\}\,\frac{\tau(z)}{1-\tau(z)}$$

Aspiration Effect. The higher tax rate in the earnings band $(z, z+dz)$ induces young workers who anticipate being in this band when old to invest less in future wage increases.

In particular, each young person in the band $(z_y(z), z_y(z+dz))$ reduces earnings by $-E_{yo} \cdot \frac{d\tau}{1-\tau(z)} \cdot z_y(z)$. Since the total number of young workers responding through this channel is given by $h_y(z_y(z)) \frac{dz_y}{dz} dz = h_o(z)\, dz$, the total tax revenue implications of the aspiration effect can be written as

$$\Delta^{AS} = -d\tau dz \cdot z_y(z)\, h_o(z) \cdot E_{yo} \cdot \frac{\tau(z_y(z))}{1-\tau(z)}$$

Accumulation Effect. The labor supply response of young workers in the band $(z, z+dz)$ affects the wage rate and earnings of those workers when they become old. This effect implies that each old person in the band $(z_o(z), z_o(z+dz))$ reduces earnings by $-E_{oy} \cdot \frac{d\tau}{1-\tau(z)} \cdot z_o(z)$. The number of old workers affected $h_o(z_o(z)) \frac{dz_o}{dz} dz = h_y(z)\, dz$, and so the total accumulation effect on tax revenue is given by

$$\Delta^{AC} = -d\tau dz \cdot z_o(z)\, h_y(z) \cdot E_{oy} \cdot \frac{\tau(z_o(z))}{1-\tau(z)}$$

Optimality. At the optimal tax schedule, the sum of the different social welfare effects derived above must be zero:

$$\Delta^M + \Delta^E + \Delta^{AS} + \Delta^{AC} = 0$$

By inserting the above effects in this optimality and noting that by the definition of $\alpha(z)$, $h_y(z) = \frac{\alpha(z)}{1-\alpha(z)} h_o(z)$, we obtain the result in Proposition 6. $\square$

The optimal age-independent tax schedule in Proposition 6 depends on weighted averages of the terms that were also present in the age-dependent tax schedules for the young and the old. Both the numerator and the denominator of equation (3.3.3) are averages of their counterparts for the age-dependent case in Proposition 4, where the weight on the young is given by the proportion of individuals at that earnings level who are young, $\alpha(z)$. Hence, the same basic effects that we discussed earlier in section 3.3.2 are still at play in the determination of age-independent taxes.

Because individuals have increasing earnings profiles over the life cycle, higher income levels will be populated to a larger degree by older workers than by younger workers, and vice versa at lower income levels, implying that $\alpha(z)$ is decreasing in z. This implies that at the bottom the optimal age-independent tax rate $\tau(z)$ puts a relatively high weight on the young and is therefore closer to the age-dependent tax rate on the young $\tau_y(z)$, whereas at the top the optimal age-independent tax rate $\tau(z)$ puts a relatively high weight on the old and is thus closer to the age-dependent tax rate on the old $\tau_o(z)$. This in turn implies that the earlier conclusions regarding optimal age-dependence (i.e., the difference between $\tau_y(z)$ and $\tau_o(z)$ at each earnings level) in the age-independent case manifest themselves as an effect on the profile of the marginal tax rate with respect to earnings (progressivity of $\tau(z)$ with respect to earnings). If there is a welfare argument for age-dependence favoring the old ($\tau_o(z) < \tau_y(z)$), this would in itself lower marginal tax rate progressivity in the age-independent schedule as higher earnings levels put more weight on $\tau_o(z)$.

3.4. How Big Are Career Effects?

Having established how the optimal way for governments to tax income depends on the career effects of work effort, the natural next question is how large these career effects actually are in practice. In sections 3.2 and 3.3 above we have shown that the key sufficient statistics for optimal income taxation are the long-run earnings elasticities (including career effects) of the two age groups to the tax rate at each age. As argued earlier, this is not what is identified by the micro literature on labor supply and taxable income responses, which mostly studies *short-run* earnings responses to *contemporaneous* tax rates. In practice, this literature comes closer to estimating the static elasticity e in our framework than the dynamic career-inclusive elasticities E_{ij} elasticities (see Piketty & Saez, 2013 for a similar argument).

Nevertheless, as equation (3.2.4) shows, the E_{ij} elasticities are functions of the underlying static elasticity e and the elasticity of future wage rates with respect to current earnings δ. While the voluminous literature on labor supply and taxable income responses can serve as a guide to what a reasonable value for the static elasticity e is, there is no such ready guidance when it comes to a reasonable value of the career elasticity δ. A careful estimation of this parameter is beyond the scope of this chapter, but there is a very large literature on experience-earnings profiles in labor economics from which we can learn something about the likely size of δ. We therefore conduct a meta-analysis of this literature, focusing on 17 empirical papers studying the effects of experience, tenure and seniority on wages whose estimates permit the derivation of an estimate of δ.

In order to derive this estimate, we must perform a simple transformation of the reported estimates as most of these papers model log wages as polynomials in experience along the lines of

$$\ln(w) = \alpha + \beta_1 EXP + \beta_2 EXP^2 + \varepsilon$$

whereas we want to estimate an elasticity $\delta = \partial \ln(w) / \partial \ln(EXP)$. To derive an estimate of δ we note that by the chain rule $\frac{\partial \ln(w)}{\partial \ln(x)} = \frac{\partial \ln(w)}{\partial x} \frac{\partial x}{\partial \ln(x)}$ and by the inverse rule of calculus $\frac{\partial x}{\partial \ln(x)} = \left[\frac{\partial \ln(x)}{\partial x}\right]^{-1} = x$ and so we can derive an estimate of δ as

$$\hat{\delta} = \left[\hat{\beta}_1 + 2\hat{\beta}_2 E\bar{X}P\right] E\bar{X}P \tag{3.4.1}$$

where $E\bar{X}P$ is the sample mean of EXP, and we can obtain standard errors by the delta method wherever the papers provide the necessary variances. We can also extend this to higher-order polynomials where the appropriate sample means are available.

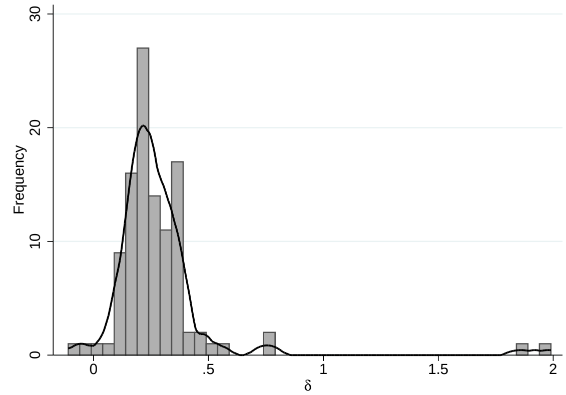
Many papers use multiple measures of experience, for example total labor market experience EXP and tenure in the individual's current job TEN as in equation (3.4.2).

$$\ln(w) = \alpha + \beta_1 EXP + \beta_2 EXP^2 + \gamma_1 TEN + \gamma_2 TEN^2 + \varepsilon \tag{3.4.2}$$

In this case, a similar derivation to that above suggests that we should use variants of $\hat{\delta} = \left[\hat{\beta}_1 + 2\hat{\beta}_2 E\bar{X}P\right] E\bar{X}P + \left[\hat{\gamma}_1 + 2\hat{\gamma}_2 T\bar{E}N\right] T\bar{E}N$ as our estimate. Finally, some papers, particularly those using more structural methods, present tables of predicted wages at various levels of experience rather than polynomials in experience. For these, we combine the estimated wage levels by simply regressing the predicted log wage on the log of experience, and again obtaining standard errors by the delta method where possible.

Applying these methods we are able to derive 108 estimates of δ. A full table of the estimates along with references to the exact locations in the papers and the methods used by the authors is in the online appendix, but Table 1 summarizes our findings. For each of the 17 papers, Table 1 presents the dataset(s) used, the population(s) studied, and the method(s) employed, as well as the average derived δ and its standard error, where the average δ is weighted by the number of observations used to estimate each δ in the chapter.

FIGURE 30. DERIVED ESTIMATES OF δ



Percentiles					mean	s.d.
10	25	50	75	90		
0.12	0.19	0.24	0.34	0.38	0.29	0.26

Notes: The 108 estimates of δ whose distribution is shown are derived using variants of equation (3.4.1) where appropriate (the vast majority of cases) and the regression of predicted wage levels on log experience levels as outlined in the text in the remaining cases. The black line is a kernel density estimate, and summary statistics of the distribution are displayed in the table below the figure.

Table 1 shows that while the estimates vary slightly from paper to paper, they mostly agree that δ lies roughly between 0.15 and 0.4 implying that a 10% increase in experience is associated with an increase in wages of between 1.5% and 4%. To reinforce this point, Figure 30 shows the distribution of all 108 estimates of δ with an overlaid kernel density alongside some summary statistics of the distribution which again show that 80% of the estimates lie between 0.19 and 0.38 with a mean of 0.29.

TABLE 12. IMPLIED DELTA FROM EXISTING ESTIMATES

Paper	Dataset(s) and period	Population	Method(s)	$\hat{\delta}$	s.e.
Borjas (1981)	1966 National Longitudinal Survey of Mature Men	Men aged 45-59	OLS	0.15	(N/A)
Abraham & Farber (1987)	PSID (SRC Subsample) 1968–1981	Non-union, male household heads aged 18–60	OLS / IV	0.24	(0.035)
Altonji & Shakotko (1987)	PSID (SRC Subsample) 1968–1981	White, male household heads aged 18–60	IV - GLS	0.35	(0.073)
Topel (1991)	PSID 1968–1983	White, male household heads aged 18–60	2-step bounding exercise	1.92	(N/A)
Filer (1993)	National Longitudinal Sample 1966-1984 & 1980 Census	Women aged 14–62 in NLS; Random Sample of Women from Census	OLS with predicted experience by sector	0.23	(0.009)
Light & Ureta (1995)	National Longitudinal Survey 1966–1984	Men born 1942–1952 during 1966–1981; Women born 1944–1954 during 1968–1984	IV-GLS potential and actual experience	0.23	(0.016)
Neal (1995)	1984–1990 Displaced Worker Surveys	Full-time, nonagricultural workers whose job was lost due to establishment closing	OLS with selection correction for job loss	0.29	(0.049)
Blau & Kahn (1997)	PSID 1980 & 1989	Full-time, nonagricultural employees aged 18-65	OLS	0.32	(0.039)
Bratsberg & Terrell (1998)	NLSY 1979–1991	Male, high school graduates not employed in agriculture or military/government	OLS; Altonji and Shaktotko (1987); Topel (1991)	0.20	(0.019)

IMPLIED DELTA FROM EXISTING ESTIMATES (CONT.)

Paper	Dataset(s) and period	Population	Method(s)	$\hat{\delta}$	s.e.
Flabbi & Ichino (2001)	HR data from Italian bank 1992–1995	Male workers	OLS	0.16	(0.007)
Dohmen (2004)	HR data from Dutch aircraft manufacturer 1987–1996	Permanent workers over 23	OLS	0.23	(0.004)
Schönberg (2007)	German Social Security Data 1975–1994; NLSY 1979–1994	Men in non-military employment	OLS	0.17	(0.004)
Altonji *et al.* (2009)	PSID (SRC and SEO Subsamples) 1978–1996	Male household heads aged 18–62	Structural	0.16	(N/A)
Grogger (2009)	Florida Family Transition Program 1994–1995	Work-contingent financial aid recipients	Selection corrected OLS	0.26	(0.101)
Regan & Oaxaca (2009)	NLSY 1979–1990; PSID 1968–1990; Integrated Public Use Microdata Sample 1990	Individuals aged 25–33 in 1990 (NLSY/IPUMS); Household heads aged 18–55 in 1990 (PSID/IPUMS)	OLS with predicted Experience	0.36	(0.051)
Yamaguchi (2009)	NLSY 1979–2004	White, male, high school and college graduates	Structural	0.26	(N/A)
Buchinsky *et al.* (2010)	PSID 1975–1992	Household heads aged 18–65 appearing $\geq$3 times	Structural	0.21	(0.017)

Notes: The $\hat{\delta}$ column shows the average of the estimates of δ derived from the estimates in the paper according to variants of equation (3.4.1) where appropriate (the vast majority of cases) and the regression of predicted wage levels on log experience levels as described in the text in the remaining cases. The estimates are weighted by the number of observations used to estimate them and where possible, the standard error of the estimate is computed using the delta method.

3.5. Numerical Simulations

3.5.1. Methodology. Our simulation method extends the procedure developed by Saez (2001) to a setting with dynamic wage rate effects. To perform numerical simulations, we first have to calibrate the three primitives of the model: the distribution of innate ability $F(n)$, the function relating the wage rate when old to innate ability and effort when young $\omega(n, z_y/n)$, and the static earnings elasticity parameter e. In the existing literature where the wage rate is exogenously given by innate ability, it is sufficient to use the first-order condition for earnings and an assumption about the earnings elasticity e to infer the ability level of an individual from the observed earnings and marginal tax rate of the individual. In our setting where the wage rate when old is endogenous to effort when young, the ability distribution cannot be determined quite so straightforwardly. Below we describe how $F(n)$ and $\omega(n, z_y/n)$ are calibrated in a manner that maintains the spirit of the method in previous work.

The calibration starts from micro data containing information about earnings, marginal tax rates and age in the United States. We obtain data on earnings and age from the 2007 round of the Panel Study of Income Dynamics (PSID), which we combine with the NBER TAXSIM model to get data on marginal tax rates.[11] To operationalize the simplification to two age groups in our model, we split the sample into the young and the old using the median age in the sample (equal to 41 years) as a cutoff. We estimate smooth earnings distributions of the young and the old from the PSID data using an adaptive kernel density estimator. Since the data are sparse for high earners and affected by top-coding of income, we follow the standard approach in the literature and fit a Pareto distribution to the upper tail of the earnings distribution. In particular, we assume that earnings are Pareto distributed above an annual income level of $150,000 for both the young and the old. To estimate the Pareto distribution's shape parameter a, we note that a Pareto distribution implies $z_m/z = a/(a-1)$ where $z_m \equiv \mathsf{E}[z_i | z_i > z]$ is defined as average earnings among those with earnings above z. Hence, the Pareto parameter a is estimated by regressing z^m/z for z between $100,000 and $150,000 on a constant, and take the estimated intercept as our estimate of $a/(a-1)$.

[11] Specifically, we use the `taxsim9` module for stata available at http://www.nber.org/~taxsim/taxsim-calc9/

Having obtained the empirical distributions of earnings and marginal tax rates of the young and the old, the calibration method proceeds in the following steps. First, using the first-order condition for earnings as old (3.2.3) and an assumption about the value of e, we can back out the wage rate ω for each old person based on information about z_o and $\tau(z_o)$. This gives us a wage rate distribution for the old, $J(\omega)$, associated with the current tax system and earnings choices. Second, we create life-cycle earnings profiles by pairing earnings observations for the old z_o with earnings observations for the young z_y in the cross-sectional data that we use. This pairing is done using a no rank-reversal assumption that corresponds to the theoretical model: it is assumed that each individual's rank in the earnings distribution when old is the same as in the earnings distribution when young, i.e. $H_o(z_o(n)) = H_y(z_y(n))$. Hence, each earnings observation for the old is linked to an earnings observation for the young according to $z_y = H_y^{-1}[H_o(z_o)]$. Notice that the optimal tax problem considered above already makes such a no rank-reversal assumption by requiring that $z_y(n), z_o(n)$ are monotonically increasing in n to guarantee implementability. What we do here is to extend the assumption to the *current* (potentially non-optimal) tax system.

Third, having obtained the variables $(z_y, z_o, \tau_y, \tau_o, \omega)$ for each individual in the sample, it is now possible to use the first-order condition for earnings as young (3.2.2) along with assumptions about the values of e and δ to infer innate ability n for every individual. For simplicity, we assume that the career elasticity δ is constant across individuals of different abilities, i.e. we assume that the wage rate when old $\omega(n, z_y/n)$ is iso-elastic with respect to effort when young z_y/n, and show simulation results for three different scenarios: a benchmark scenario with $\delta = 0$ and scenarios with career elasticities $\delta = 0.2$ and $\delta = 0.4$ in order to span the realistic range established in the meta-analysis above. Fourth, we specify that the wage rate when old is $\omega = \omega(n, z_y/n) = \omega_0(n) \cdot \left(\frac{z_y}{n}\right)^{\delta}$ where $\omega_0(n)$ is the baseline wage for an old person with innate ability n in the absence of any career investment effects. As the preceding steps have established information on n, z_y, ω for each individual and we make an assumption about δ, we can back out a baseline wage $\omega_0(n)$ ensuring that the function $\omega(\cdot)$ is satisfied for every individual. This concludes the calibration as we now have information about all the primitives of the model.

Finally, in order to simulate optimal tax rates, we must specify the social welfare criterion and the aggregate tax revenue requirement R. We follow the literature and adopt a CRRA social welfare function $\Psi[U] = U^{1-\gamma}/(1-\gamma)$, where $\gamma \geq 0$ measures preferences for equity. We consider a case with "moderate" equity preferences ($\gamma = 1$) and a case with "strong" equity preferences ($\gamma = 10$). The revenue requirement R is set equal to 10,000. The aggregate income varies from one simulation to the other as income levels are endogenous to the tax schedule, but this revenue requirement corresponds to between 8% and 11% of aggregate income. We always check that the optimal tax schedule leads to $z_y(n), z_o(n)$ that are everywhere increasing in n as this is a necessary and sufficient condition for the path of $z_y(n), z_o(n)$ to be implementable via a truthful mechanism (as described earlier).

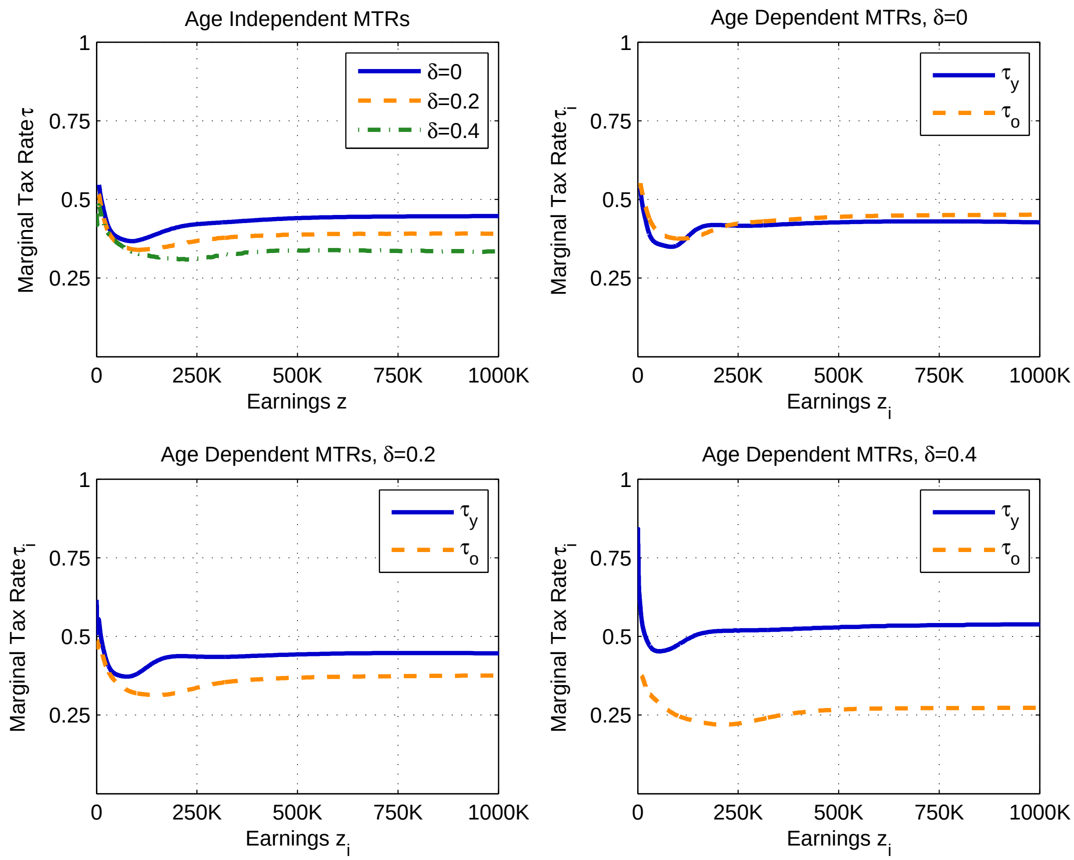
3.5.2. Results. Figures 2 and 3 show simulation results for age-independent and age-dependent tax schedules under various plausible levels of the parameters of the model. We assume that the elasticity parameter in the utility function is given by $e = 0.5$ throughout (corresponding to the static earnings elasticity without career effects), and consider three different values for the behavioral career elasticity $\delta \in \{0, 0.2, 0.4\}$ as well as two values of inequality aversion $\gamma \in \{1, 10\}$.

In Figure 2 where $\gamma = 1$, simulations of the age-independent tax schedule in the top-left panel show that when behavioral career effects of work effort are stronger, marginal tax rates are reduced everywhere. The age-independent marginal tax rate asymptotes to about 44% when $\delta = 0$, 39% when $\delta = 0.2$, and 34% when $\delta = 0.4$. We can also see that the U-shape of the optimal tax schedules becomes less pronounced as career effects become stronger, demonstrating our earlier conclusion that age-independent tax schedules exhibit less progressivity when accounting for endogenous career effects than in standard models without such effects.[12]

For the age-dependent case, when there are no behavioral career effects of work effort (so that the wage as old is determined mechanically from innate ability according to $\omega = \omega_0(n)$), the optimal tax system is associated with weak age dependence favoring the young—the marginal tax rate asymptotes to about 45% for the old, but only 43% for the young. However, once we start introducing behavioral career effects through a positive δ,

[12]Moreover, notice that due to the same effect, age-independent tax schedules also exhibit less progressivity than age-dependent tax schedules for a given strength of the career effect δ.

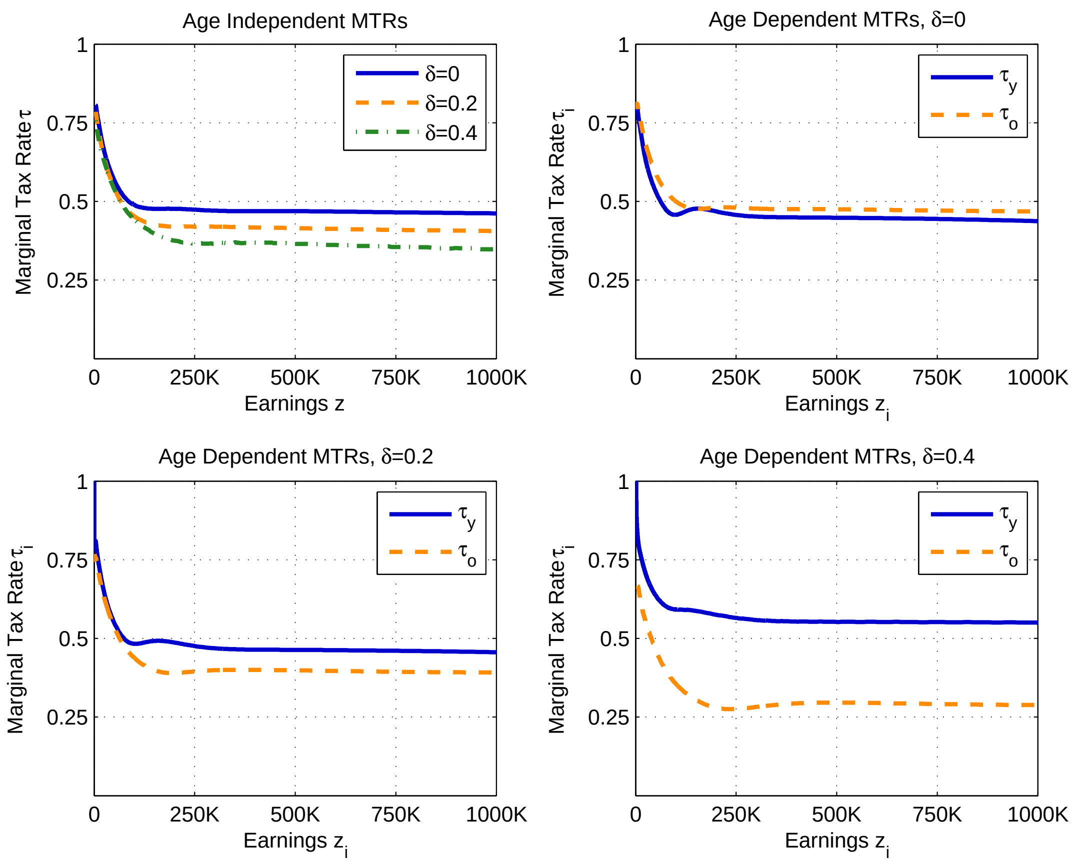
FIGURE 31. OPTIMAL MARGINAL TAX RATES, $e = 0.5$, $\gamma = 1$



optimal age dependence quickly shifts in favor of the old. At a modest behavioral career effect of $\delta = 0.2$, marginal tax rates asymptote to around 45% for the young and 37% for the old. For a stronger behavioral career effect of $\delta = 0.4$, age dependence in favor of the old becomes extremely strong with the young asymptoting to a tax rate around 53% and the old asymptoting to a tax rate around 27%.

In Figure 3 where inequality aversion is stronger at $\gamma = 10$, marginal tax rates are everywhere higher with stronger effects at the bottom than at the top of the distribution due to the fact that the social welfare weight $g(n)$ converges to zero as n becomes very large under any γ. Nevertheless, the qualitative conclusions regarding the implications of career effects for the level of marginal tax rates and age dependence remain the same. These simulations thus demonstrate that for plausible parameter values, the effects our analytical results highlight are quantitatively important with significant implications for both age-independent and age-dependent tax schedules. The stronger are behavioral

FIGURE 32. OPTIMAL MARGINAL TAX RATES, $e = 0.5$, $\gamma = 10$



career effects, the lower should be age-independent marginal tax rates and the greater should be the degree of age dependence in favor of the old.

3.6. Conclusion

The fact that wage paths are endogenous to hours worked and therefore to tax rates has been neglected in the optimal income tax literature. In this chapter, we have presented analytical results on optimal income taxation when future wages depend on current hours worked and explored the quantitative importance of such career effects through simulations using US data. In the case of age-independent annual tax schedules, career effects of work effort reduce the level of marginal tax rates at all earnings levels and make marginal tax rate schedules flatter. In the case of age-dependent annual tax schedules, career effects of work effort reduce the level of marginal tax rates on the old and increase the level of marginal tax rates on the young, conditional on earnings, thereby providing an argument for age-dependence favoring the old. Our simulations for the US suggest that reasonably modest career effects of work effort may call for substantially lower taxes on the old. We interpret these results as being driven by two effects: an elasticity effect coming from how career effects change earnings elasticities with respect to taxes at different points in the life cycle, and an equity effect coming from how career effects create a negative correlation between age and innate ability conditional on earnings. These two effects dominate an offsetting effect coming from the fact that the hazard ratio of the earnings distribution is larger among the old than among the young. These findings are opposite to the recent literature on age-dependent taxation based on the standard framework with exogenous wage paths (Kremer, 2001; Weinzierl, 2011; Golosov *et al.*, 2011; Farhi & Werning, 2012).

Our framework is highly stylized in order to highlight the implications of career effects as starkly as possible. In future work it would be interesting to relax the assumption of quasilinear utility so as to introduce savings into the model, which may interact with the optimal income tax schedule and age dependence in important ways. Also, while our analytical framework did not assume that the strength of behavioral career effects was constant throughout the ability distribution, our numerical simulations for the US were based on this simplifying assumption (i.e. we assumed that δ was constant). However, some empirical evidence suggests that experience effects on future wages are larger at the top than at the bottom of the distribution (e.g. Card & Hyslop, 2005), which has potentially important implications for the optimal progressivity of marginal tax rates. Perhaps the most important call for future research emerging from this chapter is the need to explore ways to credibly estimate earnings elasticities that incorporate dynamic wage rate effects to allow for proper implementation of the expressions for optimal income taxes derived here.

3.7. Proofs of Lemmas 6 - 8

The derivation of the earnings elasticities is a straightforward application of the implicit function theorem to the system of two first-order conditions.

$$f(z,1-\tau)=\begin{pmatrix} f_y \\ f_o \end{pmatrix}=\begin{pmatrix} 1-\tau_y(z_y)-\left(\frac{z_y}{n}\right)^{\frac{1}{e}}\frac{1}{n}+\left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y} \\ 1-\tau_o(z_o)-\left(\frac{z_o}{\omega}\right)^{\frac{1}{e}}\frac{1}{\omega} \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.7.1}$$

The implicit function theorem states that

$$\begin{aligned} D_{1-\tau}z &= -\left[D_z f(z,1-\tau)\right]^{-1} D_{1-\tau}f(z,1-\tau) \\ &= \begin{pmatrix} E_{yy}\frac{z_y}{1-\tau_y} & E_{yo}\frac{z_y}{1-\tau_o} \\ E_{oy}\frac{z_o}{1-\tau_y} & E_{oo}\frac{z_o}{1-\tau_o} \end{pmatrix} \end{aligned}$$

where $E_{ty}\equiv\frac{\partial z_t}{\partial 1-\tau_y}\frac{1-\tau_y}{z_t}$ and $E_{to}\equiv\frac{\partial z_t}{\partial 1-\tau_o}\frac{1-\tau_o}{z_t}$ are the earnings elasticities. Tedious algebra shows that

$$D_{1-\tau}z=\frac{e^2 z_y z_o}{(1-\tau_y)(1-\tau_o)\kappa}\begin{pmatrix} \frac{1}{ez_o}(1-\tau_o) & \frac{1+e}{e}\frac{\delta}{z_y}(1-\tau_o) \\ \frac{1+e}{e}\frac{\delta}{z_y}(1-\tau_o) & \frac{1-\tau_y}{ez_y}\left[1+\frac{z_o(1-\tau_o)}{z_y(1-\tau_y)}\delta(1+e)(1+\delta)\right] \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

where $\kappa=1+\delta(1-e\delta)(1+e)\frac{z_o\left[1-T_o'\right]}{z_y\left[1-T_y'\right]}$, and hence that

$$\begin{pmatrix} E_{yy} & E_{yo} \\ E_{oy} & E_{oo} \end{pmatrix}=\frac{1}{\kappa}\begin{pmatrix} e & e(1+e)\delta\frac{z_o\left[1-T_o'\right]}{z_y\left[1-T_y'\right]} \\ e(1+e)\delta & e\left[1+\delta(1+\delta)(1+e)\frac{z_o\left[1-T_o'\right]}{z_y\left[1-T_y'\right]}\right] \end{pmatrix}$$

From here, proving Lemmas 6 and 8 is straightforward. Proving Lemma 7 also requires differentiation of $\frac{z_o\left[1-T_o'\right]}{z_y\left[1-T_y'\right]}$. For this, we repeat the procedure above applying the implicit function again to the first order conditions to get that

$$\begin{aligned} D_\delta z &= \frac{e^2 z_y z_o}{(1-\tau_y)(1-\tau_o)\kappa}\begin{pmatrix} \frac{1}{ez_o}(1-\tau_o) & \frac{1+e}{e}\frac{\delta}{z_y}(1-\tau_o) \\ \frac{1+e}{e}\frac{\delta}{z_y}(1-\tau_o) & \frac{1-\tau_y}{ez_y}\left[1+\frac{z_o(1-\tau_o)}{z_y(1-\tau_y)}\delta(1+e)(1+\delta)\right] \end{pmatrix}\begin{pmatrix} \frac{z_o}{z_y}(1-\tau_o) \\ 0 \end{pmatrix} \\ &= \frac{1}{\kappa}\begin{bmatrix} ez_o\frac{1-T_o'}{1-T_y'} \\ e(1+e)z_o\frac{z_o\left[1-T_o'\right]}{z_y\left[1-T_y'\right]}\delta \end{bmatrix} \end{aligned}$$

and from here proving Lemma 7 is just further tedious algebra.

3.8. The Full Hamiltonian Method and Its Equivalence With the Direct Method

Analogously to the method of Mirrlees (1971) and Diamond (1998), we will treat lifetime utility as the state variable and the earnings levels as the control variables. Recall that lifetime utility is given by

$$U\left(z_y, z_o\right) = z_y - T_y\left(z_y\right) + z_o - T_o\left(z_o\right) - \frac{\left(\frac{z_y}{n}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} - \frac{\left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}}$$

Which has first order conditions

$$\begin{aligned} 1-\tau_y\left(z_y\right) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}}\frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y} &= 0 \\ 1-\tau_o\left(z_o\right) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}}\frac{1}{\omega} &= 0 \end{aligned}$$

Differentiating utility and letting dots denote derivatives with respect to ability

$$\dot{U}\left(n\right) = \left[1-T_y'\right]\dot{z}_y + \left[1-T_o'\right]\dot{z}_o - \left(\frac{z_y}{n}\right)^{1+\frac{1}{e}}\left[\frac{\dot{z}_y}{z_y} - \frac{1}{n}\right] - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}}\left[\frac{\dot{z}_o}{\omega} - \frac{\dot{\omega}}{\omega}\frac{z_o}{\omega}\right]$$

where $\frac{\dot{\omega}}{\omega} = \eta\frac{1}{n} + \delta\left(\frac{\dot{z}_y}{z_y} - \frac{1}{n}\right)$ and $\eta \equiv \frac{\partial\omega}{\partial n}\frac{n}{\omega}$. Using the first order conditions, this reduces to

$$\dot{U}\left(n\right) = \frac{1}{n}\left\{\left(\frac{z_y}{n}\right)^{1+\frac{1}{e}} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}\left(\eta-\delta\right)\right\}$$

3.8.1. Age-dependent Taxes. Turning first to the case of age-dependent taxes, the government's problem is to

$$\max\int_0^\infty \Psi\left[U\left(n\right)\right]f\left(n\right)dn$$

subject to

$$\begin{aligned} \int_0^\infty \left\{z_y\left(n\right) - T_y\left[z_y\left(n\right)\right] + z_o\left(n\right) - T_o\left[z_o\left(n\right)\right]\right\}f\left(n\right)dn &\leq \int_0^\infty \left[z_y\left(n\right) + z_o\left(n\right)\right]f\left(n\right)dn - R \\ 1-\tau_y\left(z_y\right) - \left(\frac{z_y}{n}\right)^{\frac{1}{e}}\frac{1}{n} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y} &= 0 \\ 1-\tau_o\left(z_o\right) - \left(\frac{z_o}{\omega}\right)^{\frac{1}{e}}\frac{1}{\omega} &= 0 \end{aligned}$$

where we use the first-order approach to substitute the first-order conditions for the complete incentive compatibility constraints.[13] Substituting out the tax system, this becomes

$$\max \int_0^\infty \Psi\left[U\left(n\right)\right] f\left(n\right) dn$$

subject to

$$\begin{aligned}\int_0^\infty \left\{ U\left(n\right) + \frac{\left(\frac{z_y(n)}{n}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} + \frac{\left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} \right\} f\left(n\right) dn &\leq \int_0^\infty \left[z_y\left(n\right) + z_o\left(n\right)\right] f\left(n\right) dn - R \\ \dot{U}\left(n\right) &= \frac{1}{n}\left\{ \left(\frac{z_y}{n}\right)^{1+\frac{1}{e}} + \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}} \left(\eta - \delta\right)\right\}\end{aligned}$$

Forming the Hamiltonian,

$$\begin{aligned} H\left(n\right) = & \left\{ \Psi\left[U\left(n\right)\right] - p\left[U\left(n\right) + \frac{\left(\frac{z_y(n)}{n}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} + \frac{\left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}} - z_y\left(n\right) - z_o\left(n\right)\right]\right\} f\left(n\right) \\ & + \mu\left(n\right)\frac{1}{n}\left\{ \left(\frac{z_y\left(n\right)}{n}\right)^{1+\frac{1}{e}} + \left(\frac{z_o\left(n\right)}{\omega\left(n\right)}\right)^{1+\frac{1}{e}}\left(\eta-\delta\right)\right\}\end{aligned}$$

When taxes are dependent on age, the planner has two control variables available, z_y and z_o which she can manipulate independently and so the optimality conditions are that

$$\begin{aligned} 0 = \frac{\partial H}{\partial z_y} &= -pf\left(n\right)\left[\frac{1}{n}\left(\frac{z_y}{n}\right)^{\frac{1}{e}} - 1 - \left(\frac{z_o\left(n\right)}{\omega\left(n\right)}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y}\right] \\ &\quad + \mu\left(n\right)\frac{1}{n}\left(1+\frac{1}{e}\right)\left\{\left(\frac{z_y}{n}\right)^{\frac{1}{e}}\frac{1}{n} - \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y}\left(\eta-\delta\right)\right\} \end{aligned} \tag{3.8.1}$$

$$\begin{aligned} 0 = \frac{\partial H}{\partial z_o} &= -pf\left(n\right)\left[\left(\frac{z_o\left(n\right)}{\omega\left(n\right)}\right)^{\frac{1}{e}}\frac{1}{\omega\left(n\right)} - 1\right] \\ &\quad + \mu\left(n\right)\frac{1}{n}\left(1+\frac{1}{e}\right)\left(\eta-\delta\right)\left(\frac{z_o\left(n\right)}{\omega\left(n\right)}\right)^{\frac{1}{e}}\frac{1}{\omega\left(n\right)} \end{aligned} \tag{3.8.2}$$

$$-\dot{\mu} = \frac{\partial H}{\partial U} = \left[\Psi'\left[U\left(n\right)\right] - p\right] f\left(n\right) \tag{3.8.3}$$

[13] In general, in dynamic moral hazard problems, the first-order approach is not always justified as sophisticated possible deviations may not be captured by the first-order conditions (the so-called "double deviation" problem). However, since in our setting the individual's problem is globally concave, we do not face these issues. For technical details on double deviations and the requirements for double deviations not to be an issue, see Kocherlakota (2004) in the context of unemployment insurance and Ábrahám & Pavoni (2009) in the context of income taxation with hidden saving.

First note that integrating equation (3.8.3), and using the transversality condition,

$$\mu(n) = \int_n^{\infty} \left[\Psi'\left[U(n)\right] - p\right] dF(n) = p \int_n^{\infty} \left[g(n) - 1\right] dF(n) \tag{3.8.4}$$

For the young, substituting the first order conditions into (**??**) we get that

$$-pf(n)\,T'_y \quad + \quad \mu\left(1+\frac{1}{e}\right)\frac{1}{n}\left\{\left[1-T'_y\right] + \left[1-T'_o\right]\frac{z_o}{z_y}\delta\left(1+\delta-\eta\right)\right\} = 0$$

which combined with equation (3.8.4) yields

$$\frac{T'_y}{1-T'_y} = \frac{\int_n^{\infty}\left[1-g(n)\right]f(n)\,dn}{nf(n)}\left(1+\frac{1}{e}\right)\left(1+\delta\frac{z_o\left[1-T'_o\right]}{z_y\left[1-T'_y\right]}\left[1+\delta-\eta\right]\right) \tag{3.8.5}$$

Turning to the old, equation (**??**) together with the first order conditions imply that

$$pf(n)\,T'_o \quad = \quad -\mu\left(1+\frac{1}{e}\right)(\eta-\delta)\frac{\left[1-T'_o\right]}{n}$$

$$\frac{T'_o}{1-T'_o} \quad = \quad \frac{\int_n^{\infty}\left[1-g(n)\right]f(n)\,dn}{nf(n)}\left(1+\frac{1}{e}\right)(\eta-\delta) \tag{3.8.6}$$

3.8.2. Age-Independent Taxes. In the case of age-independent taxes, the setup is the same as above for age-dependent taxes. However, the planner faces an additional constraint, namely that taxes be age-independent, or that $T_y(z) = T_o(z)$. To operationalize this constraint, note that it implies that $T_y(z_y) = T_o(z_o)$ whenever $z_y(n) = z_o(n')$. In particular, defining $n_o(n)$ as the ability level of the individual whose earnings when old are equal to the earnings when young of an individual of ability n, it is the case that $z_y(n) = z_o\left[n_o(n)\right]$. In effect, this constraint limits the number of control variables available to the planner to one (either $z_y(n)$ or $z_o(n)$. Without loss of generality we will work with $z_y(n)$). It further implies that changes in the control variable $z_y(n)$ are also changes in the earnings when old of individuals with ability $n_o(n)$.

This means that the optimality conditions on the Hamiltonian are instead that

$$\begin{aligned}
0 &= \frac{\partial H(n)}{\partial z_y(n)} + \frac{\partial H[n_o(n)]}{\partial z_o(n)} \\
&= -pf(n)\left[\frac{1}{n}\left(\frac{z_y(n)}{n}\right)^{\frac{1}{e}} - 1 - \left(\frac{z_o(n)}{\omega(n)}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y(n)}\right] \\
&\quad -pf[n_o(n)]\left[\left(\frac{z_o[n_o(n)]}{\omega[n_o(n)]}\right)^{\frac{1}{e}}\frac{1}{\omega[n_o(n)]} - 1\right] \\
&\quad +\mu(n)\frac{1}{n}\left(1+\frac{1}{e}\right)\left\{\left(\frac{z_y}{n}\right)^{\frac{1}{e}}\frac{1}{n} - \left(\frac{z_o}{\omega}\right)^{1+\frac{1}{e}}\frac{\delta}{z_y(n)}(\eta-\delta)\right\} \\
&\quad +\mu[n_o(n)]\frac{1}{n_o(n)}\left(1+\frac{1}{e}\right)(\eta-\delta)\left(\frac{z_o[n_o(n)]}{\omega[n_o(n)]}\right)^{\frac{1}{e}}\frac{1}{\omega[n_o(n)]} \qquad (3.8.7) \\
-\dot{\mu} = \frac{\partial H}{\partial U} &= [\Psi'[U(n)] - p]\, f(n)
\end{aligned}$$

Using the first order conditions to substitute back in the tax terms, we can rewrite equation (3.8.7) as

$$\begin{aligned}
0 &= -pf(n)\,T_y'[z_y(n)] - pf[n_o(n)]\,T_o'[z_o[n_o(n)]] \\
&\quad +\mu(n)\frac{1}{n}\left(1+\frac{1}{e}\right)\left\{1 - T_y'[z_y(n)] + (1 - T_o'[z_o(n)])\frac{z_o(n)}{z_y(n)}\delta(1+\delta-\eta)\right\} \\
&\quad +\mu[n_o(n)]\frac{1}{n_o(n)}\left(1+\frac{1}{e}\right)(\eta-\delta)(1 - T_o'[z_o[n_o(n)]])
\end{aligned}$$

or that

$$\begin{aligned}
\frac{T'[z_y(n)]}{1 - T'[z_y(n)]} &= \left\{\frac{\int_n^\infty [1-g(n)]\,f(n)\,dn}{n}\left(1 + \delta\frac{z_o(n)(1 - T'[z_o(n)])}{z_y(n)(1 - T'[z_y(n)])}[1+\delta-\eta]\right)\right. \\
&\quad \left. + \frac{\int_{n_o(n)}^\infty [1-g(n)]\,f(n)\,dn}{n_o(n)}(\eta-\delta)\right\} \times \left(1+\frac{1}{e}\right)\frac{1}{f(n)+f[n_o(n)]} \qquad (3.8.8)
\end{aligned}$$

3.8.3. Equivalence of the Hamiltonian and Direct Methods. Here we demonstrate the equivalence of the two methods for the age-dependent tax schedule for the young. The demonstration for the age-dependent tax schedule for the old and the age-independent tax schedule follow the same steps, and are left to the interested reader. The following lemma akin to Lemma 1 in Saez (2001) but for our setting will be useful in demonstrating this equivalence.

LEMMA 9. *For any tax schedule T not necessarily optimal and not necessarily age-dependent, the earnings functions z_{yn} and z_{on} are non-decreasing and satisfy the following system of differential equations.*

$$\frac{\dot{z}_y}{z_y} = \frac{1}{n}\left[1+E_{yy}+E_{yo}\eta\right]-\dot{z}_y\frac{T''_y}{1-T'_y}E_{yy}-\dot{z}_o\frac{T''_o}{1-T'_o}E_{yo} \tag{3.8.9}$$

$$\frac{\dot{z}_o}{z_o} = \frac{1}{n}\left[E_{oy}+\left(1+E_{oo}\right)\eta\right]-\dot{z}_y\frac{T''_y}{1-T'_y}E_{oy}-\dot{z}_o\frac{T''_o}{1-T'_o}E_{oo} \tag{3.8.10}$$

If equations (3.8.9) and (3.8.10) lead to $\dot{z}_y<0$ or $\dot{z}_o<0$ then z_{yn} (z_{on}) is discontinuous and (3.8.9) or (3.8.10) does not hold.

PROOF. Starting with the young, first note that $\dot{z}_{yn}/z_{yn} = \left(\dot{l}_{yn}/l_{yn}\right)+(1/n)$. Since there are no income effects, we can write the labor supply of a young individual of ability n as a function of the wages in the two periods $l_{yn}=l_y\left(w_{yn},w_{on}\right)$ where $w_{yn}=n\left(1-T'_y\right)$ and $w_{on}=\omega_n\left(1-T'_o\right)$. This means that

$$\dot{l}_{yn} = \frac{\partial l_y}{\partial w_y}\left[1-T'_y-nT''_y\dot{z}_{yn}\right]+\frac{\partial l_y}{\partial w_o}\left[\dot{\omega}_n\left(1-T'_o\right)-\omega_nT''_o\dot{z}_{on}\right]$$

Now $\dot{\omega}_n=\frac{\partial\omega_n}{\partial n}=\eta\frac{\omega_n}{n}$ where $\eta\equiv\frac{\partial\omega}{\partial n}\frac{n}{\omega}$ is the elasticity of the wage when old with respect to ability. Then, using the labor supply/earnings elasticities $E_{ij}\equiv\frac{\partial z_i}{\partial 1-T'_j}\frac{1-T'_j}{z_y}=\frac{\partial l_i}{\partial w_j}\frac{w_j}{l_i}$ we get that

$$\frac{\dot{l}_y}{l_y}=E_{yy}\frac{1}{n}-\dot{z}_y\frac{T''_y}{1-T'_y}E_{yy}+E_{yo}\frac{\eta}{n}-\dot{z}_o\frac{T''_o}{1-T'_o}E_{yo}$$

and plugging everything in and rearranging we get that

$$\frac{\dot{z}_y}{z_y}=\frac{1}{n}\left[1+E_{yy}+E_{yo}\eta\right]-\dot{z}_y\frac{T''_y}{1-T'_y}E_{yy}-\dot{z}_o\frac{T''_o}{1-T'_o}E_{yo}$$

By exactly the same reasoning, $\dot{z}_{on}/z_{on}=\left(\dot{l}_{on}/l_{on}\right)+\left(\dot{\omega}_n/\omega_n\right)$ where $l_{on}=l_o\left(w_{yn},w_{on}\right)$ so that

$$\dot{l}_{on} = \frac{\partial l_o}{\partial w_y}\left[1-T'_y-nT''_y\dot{z_{yn}}\right]+\frac{\partial l_o}{\partial w_o}\left[\dot{\omega}_n\left(1-T'_o\right)-\omega_nT''_o\dot{z}_{on}\right]$$

and

$$\frac{\dot{l}_o}{l_o}=E_{oy}\frac{1}{n}-\dot{z}_y\frac{T''_y}{1-T'_y}E_{oy}+E_{oo}\frac{\dot{\omega}}{\omega}-\dot{z}_o\frac{T''_o}{1-T'_o}E_{oo}$$

so plugging in and rearranging

$$\frac{\dot{z}_o}{z_o} = \frac{1}{n}\left[E_{oy}+\left(1+E_{oo}\right)\eta\right]-\dot{z}_y\frac{T_y''}{1-T_y'}E_{oy}-\dot{z}_o\frac{T_o''}{1-T_o'}E_{oo}$$

which finishes the proof. □

In particular, lemma 9 shows that for a tax schedule linearized around the optimum, we will have

$$\dot{z}_y = \frac{z_y}{n}\left[1+E_{yy}+E_{yo}\eta\right] \tag{3.8.11}$$

$$\dot{z}_o = \frac{z_o}{n}\left[E_{oy}+\left(1+E_{oo}\right)\eta\right] \tag{3.8.12}$$

It is also useful to note that combining (3.8.5) and (3.8.6),

$$\frac{T_o' z_o}{T_y' z_y} = \frac{\left(\eta-\delta\right)\tilde{z}}{1+\delta\tilde{z}\left(1+\delta-\eta\right)} \tag{3.8.13}$$

where we define $\tilde{z}=\frac{\left[1-T_o'\right]z_o}{\left[1-T_y'\right]z_y}$ for convenience.

The direct approach for the young gave us

$$\frac{\tau_y\left(z^*\right)}{1-\tau_y\left(z^*\right)} = \frac{\int_{z^*}^{\infty}\left(1-g_y\left(z\right)\right)dH_y\left(z\right)}{z^* h_y\left(z^*\right)\left[E_{yy}+E_{oy}\frac{\tau_o\left[z_o\left(z^*\right)\right]z_o\left(z^*\right)}{\tau_y\left(z^*\right)z^*}\right]} \tag{3.8.14}$$

Using (3.8.13) in the definition of $A_y\left(z\right)$ in equation (3.3.1),

$$A_y\left(z\right)^{-1} = \frac{1+\delta\tilde{z}\left(1+\delta-\eta\right)}{E_{yy}\left[1+\delta\tilde{z}\left(1+\delta-\eta\right)\right]+E_{oy}\left(\eta-\delta\right)\tilde{z}}$$

and using the definitions of the elasticities in (3.2.4)

$$\begin{aligned}
A_y\left(z\right)\left[1+\delta\tilde{z}\left(1+\delta-\eta\right)\right] &= E_{yy}\left[1+\delta\tilde{z}\left(1+\delta\right)\right]-E_{oy}\delta\tilde{z}+\left[E_{oy}\tilde{z}-E_{yy}\delta\tilde{z}\right]\eta \\
&= \frac{e}{\kappa}\left\{\frac{e}{1+e}+\frac{\kappa}{1+e}+e\delta\tilde{z}\eta\right\} \\
&= \frac{e}{1+e}\left[1+E_{yy}+E_{yo}\eta\right] \qquad (3.8.15)
\end{aligned}$$

Then, combining (3.8.15) with (3.8.11) and noting that by definition $h_y\left(z_y\right)\dot{z}_y = f\left[n_y\left(z_y\right)\right]$ and that $-\mu\left(n\right)/p = \int_n^{\infty}\left[1-g\left(n\right)\right]dF\left(n\right) = \int_{z_y(n)}^{\infty}\left[1-g_y\left(z\right)\right]dH_y\left(z\right)$, we demonstrate the equivalence of the Hamiltonian solution (3.8.5) and the direct solution (3.8.14) for the young. Exactly analogous steps and noting that $h_o\left(z_o\right)\dot{z}_o = f\left[n_o\left(z_o\right)\right]$

and $-\mu(n)/p = \int_n^\infty [1 - g(n)]\, dF(n) = \int_{z_o(n)}^\infty [1 - g_o(z)]\, dH_o(z)$ demonstrate the equivalence of the Hamiltonian solution (3.8.6) and the direct solution for the old. Combining these two sets of results demonstrates the equivalence for the age-dependent case.

Bibliography

Ábrahám, Árpád, & Pavoni, Nicola. 2009. *Optimal Income Taxation and Hidden Borrowing and Lending: The First-Order Approach in Two Periods*. mimeo: University College London. 167

Abraham, Katharine G., & Farber, Henry S. 1987. Job Duration, Seniority, and Earnings. *American Economic Review*, **77**, 278–297. 157

Agarwal, Sumit, Liu, Chunlin, & Souleles, Nicholas S. 2007. The Reaction of Consumer Spending and Debt to Tax Rebates – Evidence From Consumer Credit Data. *Journal of Political Economy*, **115**, 986–1019. 15, 42

Akerlof, George. 1978. The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review*, **68**, 8–19. 134

Almunia, Miguel, & Lopez-Rodriguez, David. 2013. *Firms' Responses to Tax Enforcement Strategies: Evidence from Spain*. Mimeo: University of Warwick. 96

Altonji, Joseph G., & Paxson, Christina H. 1988. Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-Offs. *Journal of Labor Economics*, **6**, 254–276. 94

Altonji, Joseph G., & Shakotko, Robert A. 1987. Do Wages Rise with Job Seniority? *Review of Economic Studies*, **54**, 437–459. 157

Altonji, Joseph G., Smith, Anthony, & Vidangos, Ivan. 2009. *Modeling Earnings Dynamics*. NBER Working Paper 14743. 158

Andreoni, James, Erard, Brian, & Feinstein, Jonathan. 1998. Tax complicance. *Journal of Economic Literature*, **36**, 818–860. 76

Andrews, Dan, Sánchez, Aida Caldera, & Joansson, Åsa. 2011. *Housing Market and Structural Policies in OECD Countries*. OECD Economics Department Working Papers No. 836. 12

Auerbach, Alan J. 1988. Capital Gains Taxation in the United States: Realizations, Revenue, and Rhetoric. *Brookings Papers on Economic Activity*, **19**, 595–638. 14, 16, 37

BANKS, JAMES, & DIAMOND, PETER. 2011. The base for Direct Taxation. *Chap. 6 of:* MIRRLEES, JAMES, ADAM, STUART, BESLEY, TIM, BLUNDELL, RICHARD, BOND, STEPHEN, CHOTE, ROBERT, GAMMIE, MALCOLM, JOHNSON, PAUL, MYLES, GARETH, & POTERBA, JAMES (eds), *Dimensions of Tax Design: The Mirrlees Review*. Oxford: Oxford University Press. 134

BEN-PORATH, YORAM. 1967. The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, **75**, 352–365. 133

BENJAMIN, JOHN D., COULSON, N. EDWARD, & YANG, SHIAWEE X. 1993. Real Estate Transfer Taxes and Property Values: The Philadelphia Story. *Journal of Real Estate Finance and Economics*, **7**, 151–157. 16

BERNHEIM, B. DOUGLAS, & RANGEL, ANTONIO. 2009. Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics*, **124**, 51–104. 77, 107

BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2011. *Pillars of Prosperity: The Political Economy of Development Clusters*. Princeton University Press. 73

BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2013. Taxation and Development. *Chap. 2, pages 51–110 of:* AUERBACH, ALAN J., CHETTY, RAJ, FELDSTEIN, MARTIN, & SAEZ, EMMANUEL (eds), *Handbook of Public Economics, Volume 5*. Elsevier. 73, 77

BESLEY, TIMOTHY, MEADS, NEIL, & SURICO, PAOLO. 2011. *The Incidence of a Transactions Tax: Evidence from a Tax Holiday*. Working Paper, London School of Economics. 16, 27, 32

BEST, MICHAEL CARLOS, BROCKMEYER, ANNE, KLEVEN, HENRIK JACOBSEN, SPINNEWIJN, JOHANNES, & WASEEM, MAZHAR. 2013. *Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan*. Mimeo: London School of Economics. 77, 92, 96

BIGIO, SAKI, & ZILBERMAN, EDUARDO. 2011. Optimal Self-Employment Income Tax Enforcement. *Journal of Public Economics*, **95**. 96

BLAU, FRANCINE D., & KAHN, LAWRENCE M. 1997. Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics*, **15**, 1–42. 157

BLUNDELL, RICHARD, & MACURDY, THOMAS. 1999. Labor Supply: A Review of Alternative Approaches. *Chap. 27, pages 1559–1695 of:* ASHENFELTER, ORLEY, & CARD, DAVID (eds), *Handbook of Labor Economics, Volume 3*. North Holland: Elsevier. 133, 135, 142

BORJAS, GEORGE J. 1981. Job Mobility and Earnings over the Life Cycle. *Industrial and Labor Relations Review*, **34**, 365–376. 157

BOVENBERG, ARY LANS, & JACOBS, BAS. 2005. Redistribution and Education Subsidies are Siamese Twins. *Journal of Public Economics*, **89**, 2005–2035. 137

BRATSBERG, BERNT, & TERRELL, DEK. 1998. Experience, Tenure, and Wage Growth of Young Black and White Men. *Journal of Human Resources*, **33**, 658–682. 157

BUCHINSKY, MOSHE, FOUGÈRE, DENIS, KRAMARZ, FRANCIS, & TCHERNIS, RUSTY. 2010. Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States. *Review of Economic Studies*, **77**, 972–1001. 158

BURMAN, LEONARD E., & RANDOLPH, WILLIAM C. 1994. Measuring Permanent Responses to Capital-Gains Tax Changes in Panel Data. *American Economic Review*, **84**, 794–809. 14, 16, 37

CABRAL, MARIKA, & HOXBY, CAROLINE. 2012. *The Hated Property Tax: Salience, Tax Rates, and Tax Revolts*. NBER Working Paper No. 18514. 77

CAMPBELL, JOHN Y., & FROOT, KENNETH A. 1994. International Experiences with Securities Transaction Taxes. *Chap. 6, pages 277–308 of:* FRANKEL, JEFFREY A. (ed), *The Internationalization of Equity Markets*. University of Chicago Press. 12

CARD, DAVID, & HYSLOP, DEAN R. 2005. Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers. *Econometrica*, **73**, 1723–1770. 164

CARD, DAVID, CHETTY, RAJ, FELDSTEIN, MARTIN, & SAEZ, EMMANUEL. 2010. *Expanding Access to Administrative Data for Research in the United States*. NSF White Paper. 74

CARRILLO, PAUL, SINGHAL, MONICA, & POMERANZ, DINA. 2013. *Tax Me if You Can: Third Party Cross-Checks and Evasion Substitution*. Mimeo: Harvard University. 74

CARRILLO, PAUL E., EMRAN, M. SHAHE, & RIVADENEIRA, ANITA. 2012. *Do Cheaters Bunch Together? Profit Taxes, Withholding Rates and Tax Evasion*. Mimeo: George Washington University. 77

CHETTY, RAJ. 2012. Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply. *Econometricsa*, **80**, 969–1018. 14, 77

CHETTY, RAJ, & SAEZ, EMMANUEL. 2013. Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics*, **5**, 1–31. 77

CHETTY, RAJ, LOONEY, ADAM, & KROFT, KORY. 2009. Salience and Taxation: Theory and Evidence. *American Economic Review*, **99**, 1145–1177. 77, 79

CHETTY, RAJ, FRIEDMAN, JOHN, OLSEN, TORE, & PISTAFERRI, LUIGI. 2011. Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *Quarterly Journal of Economics*, **126**, 749–804. 13, 14, 29, 30, 35, 73, 74, 76, 77, 79, 80, 84, 93, 94, 97, 100, 101, 109

CHETTY, RAJ, FRIEDMAN, JOHN, & SAEZ, EMMANUEL. 2013. *Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings*. forthcoming: American Economic Review. 76, 94, 102

CUNNINGHAM, CHRISTOPHER R., & ENGELHARDT, GARY V. 2008. Housing Capital-Gains Taxation and Homeowner Mobility: Evidence From the Taxpayer Relief Act of 1997. *Journal of Urban Economics*, **63**, 803–815. 16

DACHIS, BEN, DURANTON, GILLES, & TURNER, MATTHEW A. 2012. The Effects of Land Transfer Taxes on Real Estate Markets: Evidence From a Natural Experiment in Toronto. *Journal of Economic Geography*, **12**, 327–354. 16

DAHLBY, BEV. 2008. *The Marginal Cost of Public Funds: Theory and Applications*. Cambridge: MIT Press. 15, 23, 45

DE MEL, SURESH, MCKENZIE, DAVID J., & WOODRUFF, CHRISTOPHER. 2009. Measuring Microenterprise Profits: Must we Ask How the Sausage is Made? *Journal of Development Economics*, **88**. 74

DIAMOND, PETER A. 1998. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review*, **88**, 83–95. 133, 139, 144, 148, 166

DINARDO, JOHN, FORTIN, NICOLE M., & LEMIEUX, THOMAS. 1996. Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. *Econometrica*, **64**, 1001–1044. 98

DOHMEN, THOMAS J. 2004. Performance, Seniority, and Wages: Formal Salary Systems and Individual Earnings Profiles. *Labour Economics*, **11**, 741–763. 158

DUFLO, ESTHER, & SAEZ, EMMANUEL. 2003. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence From a Randomized Experiment. *Quarterly Journal of Economics*, **118**, 815–842. 77

EATON, JONATHAN, & ROSEN, HARVEY S. 1980. Taxation, Human Capital, and Uncertainty. *American Economic Review*, **70**, 705–715. 137

EUROPEAN COMMISSION. 2013. *Proposal for a Council Directive Implementing Enhanced Cooperation in the Area of Financial Transaction Tax*. http://ec.europa.eu/taxation_customs/taxation/other_taxes/financial_sector/index_en.htm. 12

FARBER, HENRY S. 1999. Mobility and Stability: The Dynamics of Job Change in Labor Markets. *Chap. 37 of:* ASHENFELTER, ORLEY, & CARD, DAVID (eds), *Handbook of Labor Economics, vol.3*. 133

FARHI, EMMANUEL, & WERNING, IVÁN. 2012. *Insurance and Taxation over the Life Cycle*. forthcoming, Review of Economic Studies. 133, 134, 164

FELDSTEIN, MARTIN, SLEMROD, JOEL, & YITZHAKI, SHLOMO. 1980. The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains. *Quarterly Journal of Economics*, **94**, 777–791. 16

FILER, RANDALL K. 1993. The Usefulness of Predicted Values for Prior Work Experience in Analyzing Labor Market Outcomes for Women. *Journal of Human Resources*, **28**, 519–537. 157

FINKELSTEIN, AMY. 2009. E-ztax: Tax Salience and Tax Rates. *Quarterly Journal of Economics*, **124**, 969–1010. 77

FISMAN, RAYMOND, & WEI, SHANG-JIN. 2004. Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China. *Journal of Political Economy*, **112**, 471–500. 74, 76

FLABBI, LUCA, & ICHINO, ANDREA. 2001. Productivity, Seniority and Wages: New Evidence from Personnel Data. *Labour Economics*, **8**, 359–387. 158

GELBER, ALEX, JONES, DAMON, & SACKS, DAN. 2013. *Earnings Adjustment to Policy: Evidence From the Social Security Earnings Test*. NBER Working Paper No. 19491. 14, 77

GERARD, FRANÇOIS, & GONZAGA, GUSTAVO. 2013. *Informal Labor and the Cost of Social Programs: Evidence from 15 Years of Unemployement Insurance in Brazil*. Mimeo: Columbia University. 74

GOLOSOV, MIKHAIL, TSYVINSKI, ALEH, & WERNING, IVAN. 2007. New Dynamic Public Finance: A User's Guide. *Chap. 5, pages 317–387 of:* ACEMOGLU, DARON, ROGOFF, KENNETH, & WOODFORD, MICHAEL (eds), *NBER Macroeconomic Annual 2006*. Cambridge: MIT Press. 133

GOLOSOV, MIKHAIL, TROSHKIN, MAXIM, & TSYVINSKI, ALEH. 2011. *Optimal Dynamic Taxes*. Mimeo: Yale University. 133, 134, 164

GOOLSBEE, AUSTAN. 2000. What Happens When You Tax the Rich? Evidence From Executive Compensation. *Journal of Political Economy*, **108**, 352–378. 14, 37

GORDON, ROGER, & LI, WEI. 2009. Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation. *Journal of Public Economics*, **93**, 855–866. 90

GORDON, ROGER H., & SLEMROD, JOEL. 2000. Are 'Real' Responses to Taxes Simply Incoe Shifting Between Corporate and Personal Tax Bases? *Pages 240–280 of:* SLEMROD, JOEL (ed), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*. New York: Russell Sage Foundation; Cambridge and London: Harvard University Press. 76, 99

GROGGER, JEFFREY. 2009. Welfare Reform, Returns to Experience, and Wages: Using Reservation Wages to Account for Sample Selection Bias. *Review of Economics and Statistics*, **91**, 490–502. 158

HILGER, NATE G. 2013. *How Does Family Income Affect College Attendance? Evidence From Timing of Parental Layoffs*. Mimeo: Brown University. 102

HOLMSTRÖM, BENGT. 1999. Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*, **66**, 169–182. 133

INTERNAL REVENUE SERVICE (IRS). 2012. *Overview of the Tax Year 2006 Tax Gap Estimates*. Washington, DC. 96

JACOBSON, LOUIS S., ALONDE, ROBERT J., & SULLIVAN, DANIEL G. 1993. Earnings Losses of Displaced Workers. *American Economic Review*, **83**, 685–709. 102

JOHNSON, DAVID S., PARKER, JONATHAN A., & SOULELES, NICHOLAS S. 2006. Household Expenditure and the Income Tax Rebates of 2001. *American Economic Review*, **96**, 1589–1610. 15, 42

JONES, DAMON. 2012. Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds. *American Economic Journal: Economic Policy*, **4**, 158–185. 77

KLEVEN, HENRIK, KNUDSEN, MARTIN, KREINER, CLAUS THUSTRUP, PEDERSEN, SOREN, & SAEZ, EMMANUEL. 2011. Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica*, **79**, 651–692. 75, 77, 96

KLEVEN, HENRIK, LANDAIS, CAMILLE, SAEZ, EMMANUEL, & SCHULTZ, ESBEN. 2013. Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax

Scheme in Denmark. *Quarterly Journal of Economics*, forthcoming. 63

KLEVEN, HENRIK JACOBSEN, & KREINER, CLAUS THUSTRUP. 2006. The Marginal Cost of Public Fundds: Hours of Work Versus Labor Force Participation. *Journal of Public Economics*, **90**, 1955–1973. 15, 23, 44

KLEVEN, HENRIK JACOBSEN, & SCHULTZ, ESBEN. 2013. *Estimating Taxable Income Responses Using Danish Tax Reforms*. American Economic Journal: Economic Policy, forthcoming. 99

KLEVEN, HENRIK JACOBSEN, & WASEEM, MAZHAR. 2013. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics*, **128**, 669–723. 13, 14, 20, 29, 30, 31, 35, 74, 77, 90, 92, 94, 99, 100, 103

KLEVEN, HENRIK JACOBSEN, KREINER, CLAUS T., & SAEZ, EMMANUEL. 2009. *Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries*. NBER Working Paper No. 15218. 27, 73, 96, 106

KOCHERLAKOTA, NARAYANA. 2004. Figuring out the Impact of Hidden Savings on Optimal Unemployment Insurance. *Review of Economic Dynamics*, **7**, 541–554. 167

KOPCZUK, WOJCIECH, & MUNROE, DAVID. 2013. *Mansion Tax: The Effect of Transfer Taxes on Residential Real Estate Market*. Mimeo: Columbia University. 16

KOPCZUK, WOJCIECH, & SLEMROD, JOEL. 2006. Putting Firms into Optimal Tax Theory. *American Economic Review Papers and Proceedings*, **96**, 130–134. 73, 107

KRAUSE, ALAN. 2009. Optimal Nonlinear Income Taxation With Learning-by-Doing. *Journal of Public Economics*, **93**, 1098–1110. 138

KREINER, CLAUS THUSTRUP, LASSEN, DAVID DREYER, & LETH-PETERSEN, SØREN. 2012. *Consumer Responses to Fiscal Stimulus Policy and Households' Cost of Liquidity*. CEPR Working Paper no. 9161. 15, 42

KREMER, MICHAEL. 2001. *Should Taxes be Independent of Age?* Mimeo: Harvard University. 134, 136, 151, 164

KUMLER, TODD, VERHOOGEN, ERIC, & FRÍAS, JUDITH. 2013. *Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico*. Mimeo: Columbia University. 74, 76, 77, 91, 92, 94

LIEBMAN, JEFFREY B., & LUTTMER, ERZO F.P. 2011. *Would People Behave Differently if They*

Better Understood Social Security? Evidence From a Field Experiment. NBER Working Paper No. 17287. 77

LIEBMAN, JEFFREY B., & ZECKHAUSER, RICHARD J. 2004. *Schmeduling*. Mimeo: Harvard Kennedy School of Government. 77, 79, 107

LIGHT, AUDREY, & URETA, MANUELITA. 1995. Early-Career Work Experience and Gender Wage Differentials. *Journal of Labor Economics*, **13**, 121–154. 157

LINCOLN INSTITUTE OF LAND POLICY. 2014. *Significant Features of the Property Tax*. http://www.lincolninst.edu/subcenters/significant-features-property-tax/Report_Real_Estate_Transfer_Charges.aspx. 12

MANNING, ALAN. 2000. Movin' On Up: Interpreting the Earnings-Esperience Profile. *Bulletin of Economic Research*, **52**, 261–295. 133

MANNING, ALAN. 2011. Imperfect Competition in the Labor Market. *Chap. 11, pages 973–1041 of:* CARD, DAVID, & ASHENFELTER, ORLEY (eds), *Handbook of Labor Economics, Volume 4b*. Elsevier. 73, 94

MATHESON, THORNTON. 2011. *Taxing Financial Transactions: Issues and Evidence*. IMF Working Paper WP/11/54. 12

MEHMOOD, AMJAD. 2012. *Annual Labour Report 2011*. Tech. rept. Labour Education Foundation, Pakistan. 91

MIAN, ATIF, & SUFI, AMIR. 2012. The Effects of Fiscal Stimulus: Evidence From the 2009 Cash For Clunkers Program. *Quarterly Journal of Economics*, **127**, 1107–1142. 14, 41

MIRRLEES, JAMES A. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, **38**, 175–208. 133, 166

MULLAINATHAN, SENDHIL. 2002. A Memory-Based Model of Bounded Rationality. *Quarterly Journal of Economics*, **117**, 735–774. 79

MULLAINATHAN, SENDHIL, SCHWARTZSTEIN, JOSHUA, & CONGDON, WILLIAM J. 2012. A Reduced-Form Approach to Behavioral Public Finance. *Annual Review of Economics*, **4**, 511–540. 77, 79

NEAL, DEREK. 1995. Industry-Specific Human Capital: Evidence from Displaced Workers. *Journal of Labor Economics*, **13**, 653–677. 157

NGAI, L. RACHEL, & TENREYRO, SILVANA. 2012. *Hot and Cold Seasons in the Housing Market*. mimeo: London School of Economics. 34

NIEHAUS, PAUL, & SUKHTANKAR, SANDIP. 2013. The Marginal Rate of Corruption in Public Programs. *Journal of Public Economics*, **104**, 52–64. 74, 77

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. 2013. *Tax Administration 2013: Comparative Information on OECD and Other Advanced and Emerging Economies*. OECD Publishing. 73

PIKETTY, THOMAS. 1997. La Redistribution Fiscale face au Chômage. *Revue Française d'Economie*, **12**, 157–201. 145

PIKETTY, THOMAS, & SAEZ, EMMANUEL. 2013. *Optimal Labor Income Taxation*. Forthcoming in Handbook of Public Economics, Vol.5. 135, 154

POMERANZ, DINA. 2013. *No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax*. Mimeo: Harvard Business School. 73, 77, 96, 106

POTERBA, JAMES M. 2002. Taxation, Risk-Taking, and Household Portfolio Behavior. *In:* AUERBACH, ALAN J., & FELDSTEIN, MARTIN (eds), *Handbook of Public Economics, vol. 3*. 12

REALLYMOVING. 2012. *Moving Costs Report September 2012*. `http://www.reallymoving.com`. 71

REGAN, TRACY L., & OAXACA, RONALD L. 2009. Work Experience as a Source of Specification Error in Earnings Models: Implications for Gender Wage Decompositions. *Journal of Population Economics*, **22**, 463–499. 158

ROGERSON, RICHARD, SHIMER, ROBERT, & WRIGHT, RANDALL. 2005. Search-Theoretic Models of the Labor Market: A Survey. *Journal of Economic Literature*, **43**, 959–988. 73, 94

ROSEN, HARVEY S. 1976. Taxes in a Labor Supply Model with Joint Wage-Hours Determination. *Econometrica*, **44**, 485–507. 94

SAEZ, EMMANUEL. 2001. Using Elasticities to Derive Optimal Tax Rates. *Review of Economic Studies*, **68**, 205–229. 133, 136, 137, 144, 145, 148, 159, 169

SAEZ, EMMANUEL. 2010. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, **2**(3). 13, 35, 74, 93, 94, 100, 127

SAEZ, EMMANUEL, SLEMROD, JOEL, & GIERTZ, SETH. 2012. The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature*, **50**, 3–50. 13, 32, 45

SCHÖNBERG, UTA. 2007. Wage Growth Due to Human Capital Accumulation and Job

Search: A Comparison between the United States and Germany. *Industrial & Labor Relations Review*, **60**, 562–586. 158

SCHWARTZSTEIN, JOSHUA. 2012. *Selective Attention and Learning*. Mimeo: Dartmouth College. 79

SHAN, HUI. 2011. The Effect of Capital Gains Taxation on Home Sales: Evidence From the Taxpayer Relief Act of 1997. *Journal of Public Economics*, **95**, 177–188. 16

SHAPIRO, MATTHEW D., & SLEMROD, JOEL. 2003a. Consumer Responses to Tax Rebates. *American Economic Review*, **93**, 381–396. 15, 42

SHAPIRO, MATTHEW D., & SLEMROD, JOEL. 2003b. Did the 2001 Tax Rebate Stimulate Spending? Evidence from Taxpayer Surveys. *Tax Policy & The Economy*, **17**, 83–109. 15, 42

SIMS, CHRISOPHER. 2003. Implications of Rational Inattention. *Journal of Monetary Economics*, **50**, 665–690. 79

SINIAVSKAIA, NATALIA. 2008. *Spending Patterns of Home Buyers*. Report for HousingEconomics.com. 42

SLEMROD, JOEL. 2007. Cheating Ourselves: The Economics of Tax Evasion. *Journal of Economic Perspectives*, **21**, 25–48. 76

SLEMROD, JOEL. 2008. Does it Matter Who Writes the Check to the Government? The Economics of Tax Remittance. *National Tax Journal*, **61**, 251–275. 73, 91

SLEMROD, JOEL, & WEBER, CAROLINE. 2012. Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, **19**(1), 25–53. 76

SLEMROD, JOEL, & YITZHAKI, SHLOMO. 2002. Tax avoidance, evasion, and administration. *Chap. 22, pages 1423–1470 of:* AUERBACH, A. J., & FELDSTEIN, M. (eds), *Handbook of Public Economics*. Handbook of Public Economics, vol. 3. Elsevier. 76

SLEMROD, JOEL, BLUMENTHAL, MARSHA, & CHRISTIAN, CHARLES. 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, **79**(3), 455–483. 77

SLEMROD, JOEL, WEBER, CAROLINE, & SHAN, HUI. 2012. *The Lock-In Effect of Housing Transfer Taxes: Evidence From a Notched Change in D.C. Policy*. mimeo: University of Michigan. 16

SPINNEWIJN, JOHANNES. 2013a. *Heterogeneity, Demand for Insurance and Adverse Selection*. Mimeo: London School of Economics. 77

SPINNEWIJN, JOHANNES. 2013b. *Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs*. Journal of the European Economic Association, forthcoming. 77

STIGLITZ, JOSEPH. 1982. Self-selection and pareto efficient taxation. *Journal of Public Economics*, **17**, 213–240. 138

TOPEL, ROBERT. 1991. Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority. *Journal of Political Economy*, **99**, 145–176. 157

VAN OMMEREN, JOS, & VAN LEUVENSTEIJN, MICHIEL. 2005. New Evidence of the Effect of Transaction Costs on Residential Mobility. *Journal of Regional Science*, **45**, 681–702. 16

WEINZIERL, MATTHEW C. 2011. The Surprising Power of Age-Dependent Taxes. *Review of Economic Studies*, **78**, 1490–1518. 134, 136, 150, 151, 164

YAMAGUCHI, SHINTARO. 2009. The Effect of Match Quality and Specific Experience on Career Decisions and Wage Growth. *Labour Economics*, **17**, 407–423. 158

YANIV, GIDEON. 1988. Withholding and Non-Withheld Tax Evasion. *Journal of Public Economics*, **35**, 183–204. 77, 95

ZILLOW.COM. 2012. *Mover Study*. `http://www.zillowblog.com`. 42

ZUCMAN, GABRIEL. 2013. The Missing Wealth of Nations: Are Europe and the U.S. net Debtors or net Creditors? *Quarterly Journal of Economics*, **128**, 1321–1364. 77