

Optimal deductibility: Theory, and evidence from a bunching decomposition*

Steven Hamilton[†]

May 10, 2018

Abstract

I define a new tax instrument, the ‘deductibility rate’, which specifies the proportion of eligible expenses a taxpayer may deduct when preparing her taxes. If the utilities of gross income and deductions are separable, then the deduction elasticity reflects the revenue leakage caused by greater deductibility. To identify this elasticity, I develop the first method to decompose bunching in taxable income into its constituent parts, exploiting the removal of a notch in the tax schedule. This setting also generates an observed counterfactual density, obviating the parametric assumptions routinely made in bunching studies. Applying this method to new administrative tax data from Australia, I find that while deductions account for just 5% of taxable income, they account for 35% of the response of taxable income to the tax rate. Based on an elasticity of taxable income of 0.06, the deduction elasticity is -0.45 , and the gross-income elasticity is 0.04. Consistent with standard optimal-tax logic, the sensitivity of deductions to the tax rate suggests that restricting deductions could raise welfare.

*My thanks to Traviss Cassidy, John DiNardo, Jim Hines, Joel Slemrod, Jeff Smith, Luke Spreen, Caroline Weber, Justin Wolfers, and seminar participants at the University of Michigan, George Washington University, the University of Technology Sydney, Monash University, the Grattan Institute, and the 2016 and 2017 National Tax Association Annual Conferences.

[†]University of Michigan. Email: steveham@umich.edu. The latest version can be downloaded [here](#).

1 Introduction

The modern literature on the behavioural response to income taxes has focused almost exclusively on the bottom line of the tax return, taxable income finding it responds fairly modestly to a change in the tax rate. I develop a new method to decompose the taxable-income response associated with bunching near discontinuities in the tax schedule into the responses available, gross income and deductions, the two principal components of taxable income. Using new administrative tax records from Australia, I find that the effect of the tax rate on deductions is an order of magnitude larger than that on gross income. The logic that the optimal tax rate is inversely related to the behavioural response suggests the anatomy of the response can be informative to policy. Namely, the large observed response indicates that limiting the ability of taxpayers to claim deductions could raise welfare.

I formalise this logic in a simple Ramsey (1927) model of optimal deductibility, in which? instead of consuming commodities, the taxpayer reports gross income and deductions in her tax return. In addition to the tax rate, the government selects the proportion of expenses that are deductible (the deductibility rate). While the elasticity of taxable income (ETI) continues to be a sufficient statistic for the deadweight loss of the tax rate, it is not sufficient for the deadweight loss of the deductibility rate. Because the latter depends on the response of taxable income to the deductibility rate, the lack of observed variation in the deductibility rate presents a challenge for empirical work.

It is more common to observe variation in the tax rate, but this won't in general induce the same response as a change in the deductibility rate. Under quasilinear, isoelastic, and separable utility, which combines the functional form used in bunching studies with the assumption often made in Ramsey models that the cross-price elasticities are zero, the two changes have equivalent effects on deductions. In that case, the response of deductions to a change in the tax rate, in the form of the deduction elasticity, is a sufficient statistic for the deadweight loss of the deductibility rate. The validity of this assumption depends on the substitutability or complementarity of gross income and deductions in practice. As in the standard Ramsey case, if they are complementary or substitutable, then the optimal deductibility rate will still depend inversely on the deduction response, but this will be either attenuated or accentuated by the gross-income response.

Bunching methods have been widely used to estimate the response of taxable income to a change in the tax rate, but they have not been used to decompose the response. I extend the standard bunching model to include gross income and deductions, and exploit the taxpayer's optimality conditions before and after bunching to derive a simple formula for the deduction elasticity, which depends on the ETI and the relative proportional changes of deductions and taxable income in bunching. To estimate the deduction elasticity using this formula, it is necessary to observe a set of taxpayers who face the discontinuity, so have an incentive to

bunch, and a set who do not.

I consider a 16% sample of Australian administrative income tax records, which to date has seen little use by academic researchers. I study the effects of a provision under which an additional 1% tax is paid by taxpayers without dependents who do not have private health insurance coverage, and whose taxable income exceeds AU\$50,000. This is a notch, in public-finance parlance, because tax liability jumps by \$500 at the threshold. This generates a strong incentive for certain taxpayers to reduce their taxable incomes to below the threshold, which requires them to decrease gross income or increase deductions. In 2009, the government raised the threshold to \$70,000, which generates a treatment group (those near the \$50,000 threshold in 2008) and a comparison group (those in the same region in 2009) whose tax returns can be compared.

I identify the range of taxpayers affected by the policy (the manipulation region) by comparing the densities of the treated and nontreated groups. Considering all taxpayers in the manipulation region with and without the treatment avoids bias due to selection into bunching, because all those who bunch and who don't bunch are always included. A simple comparison of the treated and nontreated means is, however, subject to a different selection bias because, at a given taxable income, gross income and deductions vary year-to-year in the absence of treatment. To address this, I implement a difference-in-differences design, exploiting a placebo group in the region below and adjacent to the manipulation region. These taxpayers are comparable to the treated but never receive the treatment. I address the possibility of non-parallel pre-trends by predicting the counterfactual outcome in the nontreatment period based on apparently linear pre-trends, and use these in place of the observed nontreatment outcomes.

I find that, in absolute terms, the average response in deductions accounts for around a third (\$187) of the average response in taxable income (\$527), with gross income accounting for the remaining two thirds (\$340). But because deductions constitute only 5% (\$2,380) of taxable income in the absence of treatment (\$50,535), their proportional response is an order of magnitude greater than that of gross income. For every 1% decrease in taxable income, deductions increase by 7.5%, while gross income decreases only by 0.6%. Given an estimated ETI of 0.06, this translates to a deduction elasticity of -0.45 , and a gross-income elasticity of 0.04.

Based on the optimal-tax formulas I derive in the paper, and given reasonable parameter values, each \$1 of deductions would have to generate 68¢ in external benefits, over and above the benefits to the taxpayer claiming it, in order for it to be optimal to allow taxpayers to fully deduct their expenses. If the external benefit of deductions were even as high as 30¢, the optimal deductibility rate would be just 34%. These results are driven by the large observed response of deductions to a change in the tax rate. They reflect the standard logic of the Ramsey inverse-elasticity rule that goods with high elasticities should be taxed less.

This paper informs three strands of the public-finance literature. The first is the literature on the behavioural response to income taxes, which, following Feldstein (1999), has come to focus on the ETI. Saez, Slemrod and Giertz (2012) offer a useful review of this literature. Slemrod (1998), Chetty (2009), and Doerrenberg, Peichl and Siegloch (2015) propose conditions under which the ETI is not a sufficient statistic for the welfare impact of a change in the tax rate. I follow them by proposing conditions under which the ETI is not a sufficient statistic for the optimal setting of a different tax instrument. In that regard, my work is related to that of Slemrod and Kopczuk (2002), who consider the ability of the government to set the optimal tax base as do I with the deductibility rate.

The second strand is the literature on empirical bunching methods. While the ETI literature historically used panel data to observe variation over time following tax reforms (Feldstein, 1995; Gruber and Saez, 2002), identification under these methods has been questioned (Weber, 2014). More recently, scholars have relied on bunching to identify the ETI, which is seen to offer more credible identification (Saez, 2010; Chetty, Friedman, Olsen and Pistaferri, 2011; Kleven and Waseem, 2013). Kleven (2016) offers a useful review of this literature. I extend this literature by proposing a method to decompose the bunching response.

The last strand is the developing literature on deductions, to which my main results contribute. Doerrenberg *et al.* (2015) show that the response of deductions to the tax rate is a necessary statistic for the optimal tax rate when deductions generate external benefits, and apply the traditional panel-data methods to estimate the deduction elasticity in Germany. Based on a higher ETI of 0.6, they find a deduction elasticity of -0.9 .¹ Also using German data, Schächtele (2016) notes that, when considering a measure of taxable income that excludes deductions, there is no bunching at a particular kink in the German tax schedule, which suggests that those without deductions are unresponsive to the tax rate. Paetzold (2017) estimates the deduction elasticity by applying a regression-kink design to a change in the probability of claiming deductions at a discontinuity in the marginal tax rate in Austria. Given an ETI of 0.1, he finds a deduction elasticity of -0.6 . A comparable decomposition for firms has been considered by Best, Brockmeyer, Kleven, Spinnewijn and Waseem (2015) and Bachas and Soto (2017). The methods I develop in this paper are easily adaptable to the corporate taxation setting.

In section 2, I present a model of optimal deductibility, setting out the conditions under which the deduction elasticity is a sufficient statistic for the optimal deductibility rate. In section 3, I present a bunching decomposition method, which I later rely on to estimate the deduction elasticity. In section 4, I describe the Australian tax system, the policy that I consider, and the data on which I rely. In section 5, I estimate the deduction elasticity. In

¹In the present case, if the ETI were 0.6, then the measured deduction elasticity would be -4.52 .

section 6, I conclude.

2 A Ramsey model of optimal deductibility

I develop a model of optimal deductibility to determine the conditions under which the observed response of deductions to a change in the tax rate, which I observe, is informative to policy. In doing so, I am motivated by two questions. How does the optimal deductibility rate depend on the responsiveness of deductions to the deductibility rate? And, as the deductibility rate seldom varies in practice, under what conditions can variation in the tax rate empirically be relied upon instead?

I apply the Ramsey (1927) model of optimal commodity taxation to a taxpayer who reports gross income and deductions in her tax return. I represent the tax return similarly to Feldstein (1999), but he assumes the government chooses only the tax rate applicable to taxable income. I allow the government to choose also the proportion of eligible expenditures that is deductible (the deductibility rate). The tax return can be disaggregated in many ways, but gross income and deductions are of interest to a developing literature.² The expenses are assumed to have some external value, which is why the government might permit their deductibility. In the empirical exercise later in the paper, I rely on the values of gross income and deductions reported in the tax return, so the correct interpretation of the model is that the *reporting* of the expenses generates external value. If the observed expenses data are misreported, then their true external value may be less than is represented in the model.

The taxpayer chooses consumption, c , gross income, y , and deductions, d to maximise utility, $u(c, y, d)$, subject to her budget constraint. She does so subject to the constant marginal tax rate, τ , which applies to taxable income, $z = y - \rho d$, with ρ the deductibility rate. Generating gross income reduces utility at an increasing rate ($u_y < 0$ and $u_{y,y} > 0$), while deductions increase utility at a decreasing rate ($u_d > 0$ and $u_{d,d} < 0$). This is a reduced-form representation of utility, similar to that of Feldstein (1999), in which gross income and deductions enter the utility function individually. Income generation involves disutility, for example due to the supply of labour. Deductible expenses often have consumption value, for example charitable contributions, mortgage payments, medical expenses, health-insurance premiums, or certain work-related expenses. But any portion of deductible expenses undertaken solely to generate income, so that it has no consumption value and is reported truthfully, cannot be represented as deductions in this model.³

²This literature includes Doerrenberg *et al.* (2015), Schächtele (2016), and Paetzold (2017) for personal taxation, but also Best *et al.* (2015) and Bachas and Soto (2017) for corporate taxation.

³In a standard model of optimal profit taxation, these kinds of expenses optimally will be fully deductible, with any limit on deductibility introducing a productive inefficiency. Best *et al.* (2015) note, however, that the revenue leakage caused by the misreporting of business expenses reduces welfare, meaning that less-than-full deductibility could be optimal.

Given the taxpayer's optimal gross income and deductions, indirect utility is:

$$v(\tau, \rho) = \max_{c, y, d} u(c, y, d) + \lambda^c [y - d - \tau \cdot (y - \rho d) - c],$$

with her first-order conditions yielding:

$$\frac{u_y}{u_d} = -\frac{1 - \tau}{1 - \rho\tau}.$$

As shown in figure 1, under full deductibility the tax rate does not affect the relative prices of the items, but a change in deductibility does.

The government chooses τ and ρ to maximise social welfare, which includes indirect utility and the external value of deductions, $\Phi(d)$, subject to a revenue requirement, R :

$$\max_{\tau, \rho} w(\tau, \rho) = v(\tau, \rho) + \Phi(d) + \lambda^g \cdot [\tau \cdot (y - \rho d) - R].$$

The government's first-order conditions are:

$$\frac{\partial w}{\partial \tau} = \left(1 - \frac{u_c}{\lambda^g}\right) \cdot (y - \rho d) + \frac{\Phi'(d)}{\lambda^g} \cdot \frac{\partial d}{\partial \tau} + \tau \cdot \left(\frac{\partial y}{\partial \tau} - \rho \cdot \frac{\partial d}{\partial \tau}\right) = 0 \quad (1)$$

$$\frac{\partial w}{\partial \rho} = -\left(1 - \frac{u_c}{\lambda^g}\right) \cdot \tau d + \frac{\Phi'(d)}{\lambda^g} \cdot \frac{\partial d}{\partial \rho} + \tau \cdot \left(\frac{\partial y}{\partial \rho} - \rho \cdot \frac{\partial d}{\partial \rho}\right) = 0. \quad (2)$$

The first term in equations 1 and 2 represents a transfer of funds between taxpayer and government, and the last term represents a behavioural distortion.⁴

The distortion in equation 2 is due to a change in the deductibility rate, which presents a challenge for empirical work because this seldom is observed. Because we instead observe variation in the tax rate, it is useful to consider the conditions under which such variation can be relied upon for the normative analysis of deductibility. Using the implicit function theorem, the response of deductions to the deductibility rate can be written as:

$$\frac{\partial d}{\partial \rho} = -\frac{\tau}{1 - \rho} \cdot \left[\frac{\partial d}{\partial \tau} - \left(\frac{u_{dy} - u_{yy}}{u_{dd} - u_{yd}} \right) \cdot \frac{\partial y}{\partial \tau} \right]. \quad (3)$$

As can be seen in equation 3, while the ETI is a sufficient statistic for the deadweight loss of the tax rate, as per Feldstein (1999), in general it is neither necessary nor sufficient for the deadweight loss of the deductibility rate. When variation is observed only in the tax rate, one needs to know also the second-order terms in the taxpayer's utility function. Without a functional-form assumption, it is unclear what the effect on deductions of a change in the tax

⁴When deductions generate an external benefit, the elasticity of deductions with respect to the net-of-tax rate is a necessary statistic for determining the optimal tax rate, as previously observed by Doerrenberg *et al.* (2015).

rate implies for optimal deductibility.

Modern estimation of the ETI exploits bunching around discontinuities in the tax schedule (Saez, 2010; Chetty *et al.*, 2011; Kleven and Waseem, 2013). In these studies, utility takes a particular quasilinear and isoelastic form. Analogous to that approach, I assume that functional form applies to gross income and deductions separately:⁵

$$u(c, y, d) = c - \frac{n_y}{1 + 1/e_y} \cdot \left(\frac{y}{n_y}\right)^{1+1/e_y} + \frac{n_d}{1 + 1/e_d} \cdot \left(\frac{d}{n_d}\right)^{1+1/e_d}, \quad (4)$$

in which n_y and n_d are gross income and deductions in the absence of taxes, and e_y and e_d are elasticities. The optimal gross income and deductions are then:

$$y_0 = n_y \cdot (1 - \tau)^{e_y} \quad d_0 = n_d \cdot (1 - \rho\tau)^{e_d}, \quad (5)$$

with elasticities:

$$e_y = \frac{dy}{d(1 - \tau)} \cdot \frac{1 - \tau}{y} \quad e_d = \frac{dd}{d(1 - \rho\tau)} \cdot \frac{1 - \rho\tau}{d}.$$

The effects of the tax and deductibility rates on deductions are proportional to one another because all that matters for deductions is the effective net-of-tax rate, $1 - \rho\tau$. Accordingly, equation 3 becomes:

$$\frac{\partial d}{\partial \rho} = \frac{\tau}{\rho} \cdot \frac{\partial d}{\partial \tau}. \quad (6)$$

This is illustrated in figure 1. A one-unit change in the tax rate has the same effect on deductions as a τ/ρ -unit change in the deductibility rate, but the change in the deductibility rate has no effect on gross income. Under quasilinear, isoelastic, and separable utility, the deduction elasticity is a sufficient statistic for the deadweight loss of the deductibility rate.

An increase in deductions can be achieved by increasing either of the tax or deductibility rates because both increase the implicit subsidy to deductions. This raises a question as to their relative social costs in doing so. Given the government's first-order conditions and the functional form assumption, the net effect on welfare of simultaneously raising the tax rate by one unit and lowering the deductibility rate by τ/ρ units (which leaves deductions unchanged) is:

$$\frac{\partial w}{\partial \tau} - \frac{\rho}{\tau} \cdot \frac{\partial w}{\partial \rho} = -(1 - \lambda^g) \cdot y - \lambda^g \cdot \frac{\tau}{1 - \tau} \cdot e_y y. \quad (7)$$

⁵Only quasilinearity and separability are necessary to obtain the desired mapping; isoelasticity is necessary to obtain explicit solutions for the optimal deductibility rate.

The two changes on net generate revenue at the cost of a behavioural response, with deductions unaffected. The deductibility rate should be chosen to maximise the net social benefit of deductions, and the two rates may then be adjusted together based on the impact on gross income.

If the deductibility rate has been set optimally, then the optimal tax rate is:

$$\tau^*(p^*) = \frac{1 - \lambda^g}{1 - \lambda^g \cdot (1 + e_y)}, \quad (8)$$

which takes the usual Ramsey (1927) inverse-elasticity form with the gross-income elasticity a sufficient statistic for the optimal tax rate. As a function of the prevailing tax rate, the optimal deductibility rate is:

$$\rho^*(\tau) = \frac{1}{\tau} \cdot \frac{1 - \lambda^g - \Phi'(d) \cdot e_d}{1 - \lambda^g - \lambda^g \cdot e_d}, \quad (9)$$

which is explicit only under a constant marginal external benefit of deductions.

Only if a dollar in the hands of the government is worth the same as a dollar in the hands of the taxpayer ($\lambda^g = 1$) is the deduction elasticity irrelevant.⁶ In that case, the only concern is correcting the deduction ‘externality’, so the effective deduction subsidy rate, $\rho\tau$, should be set equal to the marginal external benefit of deductions, $\Phi'(d)$, as per a Pigouvian subsidy.⁷ To the extent that government revenue is raised at a social cost, which surely is the case in practice, the deduction elasticity is a necessary statistic for the optimal deductibility rate. As long as the marginal external benefit of deductions exceeds the social marginal utility cost of public funds ($1 < \lambda^g < \Phi'(d)$), the Ramsey (1927) inverse-elasticity rule applies: the more elastic are deductions with respect to the net-of-tax rate, the lower is the optimal deductibility rate.

The functional-form assumption rules out income effects for gross income and deductions, heterogeneity in the gross-income and deduction elasticities (though this can be accommodated), adjustment frictions, and any dependence of gross income on the deductibility rate. It will be necessary to assume at least quasilinearity and isoelasticity to perform the bunching analysis later on. The separability assumption sets to zero the cross-price elasticities of gross income and deductions, as is sometimes assumed in order to derive the ‘inverse-elasticity’ representation of the Ramsey (1927) optimal tax rate. When this doesn’t hold, the optimal

⁶The term, λ^g is the ‘social marginal utility cost of public funds’, and captures the welfare loss associated with the marginal dollar raised by government given that revenue is raised at a social cost.

⁷Note also the relevance of this setting to the ‘double dividend’ debate in optimal environmental taxation (Bovenberg and de Mooij, 1994). If expenses instead generated a negative externality, then the model would stipulate a *negative* deductibility rate, implying the expenses attract an additional tax burden on top of gross income. Because deductions and gross income are separable, this tax would have no effect on gross income. This assumption implies a double dividend: the negative-externality-generating expenses could be taxed with no distortion to gross income, with the proceeds used to fund a reduction in distortive income taxation.

tax rate continues to depend inversely on the own-price elasticity, but depends also on the cross-price elasticities. To the extent that substitution to other taxed goods reduces revenue leakage, the optimal tax rate is higher, and the opposite is true for complements. Any degree of complementarity or substitutability of gross income and deductions will accentuate or attenuate the effect of the deductibility rate on welfare that I derive under separability.⁸

3 A bunching decomposition method

When a discontinuous increase in tax liability is introduced into the tax schedule, taxpayers just above the discontinuity have an incentive to reduce their taxable incomes and bunch just below it. I identify the deduction and gross-income elasticities by attributing the reduction in taxable income associated with bunching to changes in deductions and gross income. I extend the standard bunching model (Saez, 2010; Chetty *et al.*, 2011; Kleven and Waseem, 2013) to include deductions and gross income, and then derive their elasticities as functions of the ETI and their proportional changes relative to income in bunching.

I assume there is full deductibility ($\rho = 1$), and taxable income is gross income less deductions, $z = y - d$. In the absence of taxes, deductions are $n_y \sim F_d(\cdot)$ and gross income is $n_y \sim F_y(\cdot)$, which are the taxpayer's 'types'. Given her tax liability, $T(z)$, the taxpayer chooses gross income and deductions to maximise the quasilinear, isoelastic, and separable utility function specified earlier:

$$u(y, d) = y - d - T(y - d) - \frac{n_y}{1 + 1/e_y} \cdot \left(\frac{y}{n_y} \right)^{1 + 1/e_y} + \frac{n_d}{1 + 1/e_d} \cdot \left(\frac{d}{n_d} \right)^{1 + 1/e_d}.$$

A discontinuity (the treatment) is introduced into a previously linear tax schedule (the counterfactual). Deductions and gross income, and thus taxable income, are the outcomes of treatment or nontreatment. Accordingly, I use potential-outcomes notation from the program-evaluation literature (Rubin, 1974; Rosenbaum and Rubin, 1983). Deductions under treatment and nontreatment are d_1 and d_0 . I refer explicitly only to deductions, but analogous representations of taxable income and gross income are implied. For now, think of the treatment as being randomly assigned.

Under a linear tax, $T(z) = t \cdot z$, deductions are $d_0 = n_d \cdot (1 - t)^{e_d}$. Given the type distributions, this generates a deductions distribution, $h_d(d_0)$. I restrict the type distributions only insofar as they generate a smooth taxable-income distribution, $h_z(z_0) = \int h_{y,d}(z_0 -$

⁸The standard taxable-income representation (Saez *et al.*, 2012) includes a separability assumption implicitly because taxable income in a given period is itself one of several possible 'items' upon which a taxpayer could report. Taxpayers might reclassify personal income as business income, or shift income between periods in response to relative tax rates. As noted by Slemrod (1998), the taxpayer's ability to substitute into these other items undermines the sufficiency of the ETI, just as in the present case any substitution into gross income would undermine the sufficiency of the deduction elasticity.

$d_0, d_0) dd_0$. Consider the introduction of a discontinuous increase in tax liability (a notch), at $z = z^*$. The notched tax function is $T(z) = t \cdot z + \Delta t \cdot z \cdot \mathbf{1}[z > z^*]$. For a taxpayer who was located just above the threshold, the introduction of the notch will mean that reducing taxable income increases after-tax income as she avoids paying the additional tax. And the only way for her to do so is to decrease gross income or increase deductions.

The taxable income of a taxpayer who is indifferent to bunching at the threshold is $z^* + \Delta z^*$. Unlike the case of taxable income with heterogeneous elasticities, this does not define a unique buncher, but rather a set of bunchers for whom $y_0 - d_0 = z^* + \Delta z^*$. If taxpayers have the same elasticities and there are no adjustment frictions, then all taxpayers with $z_0 \in (z^*, z^* + \Delta z^*]$ will bunch at the threshold. The decision to bunch is discrete, with the taxpayer comparing her utility when she bunches to that when she doesn't bunch.

The task is to identify the deduction elasticity:

$$e_d = \frac{dd/d}{d(1-\tau)/(1-\tau)}.$$

The ETI is $e = (dz/z)/(d(1-\tau)/(1-\tau))$. The decision problem of a buncher is:

$$\max_d u(z^* - d, d) = (1-t) \cdot z^* - \frac{n_y}{1+1/e_y} \cdot \left(\frac{z^* - d}{n_y}\right)^{1+1/e_y} + \frac{n_d}{1+1/e_d} \cdot \left(\frac{d}{n_d}\right)^{1+1/e_d},$$

that is, the choice of one of the items is residual.

To derive the deduction and gross-income elasticities, I rely on the taxpayer's always equalising the marginal utilities of gross income and deductions. For bunchers, with the notch in place:

$$\left(\frac{y_1}{n_y}\right)^{1/e_y} = \left(\frac{d_1}{n_d}\right)^{1/e_d}, \quad (10)$$

with $y_1 - d_1 = z^*$, while without the notch, gross income and deductions are:

$$y_0 = n_y \cdot (1-t)^{e_y} \quad d_0 = n_d \cdot (1-t)^{e_d}. \quad (11)$$

Rearranging equations 11, substituting for n_y and n_d in equation 10, and solving for the ratio of elasticities yields:

$$\frac{e_y}{e_d} = \frac{\ln y_1 - \ln y_0}{\ln d_1 - \ln d_0}. \quad (12)$$

The ETI is the average of the deduction and gross-income elasticities weighted by the propor-

tions of taxable income for which they account:⁹

$$e = \frac{y}{z} \cdot e_y - \frac{d}{z} \cdot e_d. \quad (13)$$

Evaluating equation 13 in the absence of the notch, solving for the ratio of the gross-income and deduction elasticities, and substituting into equation 12 yields the deduction elasticity:

$$e_d = e \cdot \frac{\ln\left(\frac{\Delta d}{d_0} + 1\right)}{\frac{y_0}{z_0} \cdot \ln\left(\frac{\Delta y}{y_0} + 1\right) - \frac{d_0}{z_0} \cdot \ln\left(\frac{\Delta d}{d_0} + 1\right)} \approx e \cdot \frac{\Delta d}{\Delta z} \cdot \frac{z}{d}, \quad (14)$$

where the approximate form can be verified by the chain rule, and $\Delta d = d_1 - d_0$. Given the ETI, the deduction elasticity can be estimated via the change in deductions associated with a change in taxable income in bunching. The approach is depicted in figure 1a.

A helpful feature of equation 14 is that, outside of the ETI term, the tax-rate change is irrelevant to the deduction elasticity. For a notch, it is the *average* tax rate that changes at the threshold, but the desired elasticities are with respect to the net-of-*marginal*-tax rate. The challenge this poses has been addressed for estimating the ETI, eased by the fact that the standard bunching method exploits the taxable income responses of a set of taxpayers with the same taxable income.¹⁰ But, as I will describe in a moment, estimation of the deduction elasticity in equation 14 relies on the average response of a set of taxpayers among whom the proximity to the threshold varies. Their implicit marginal tax rates differ because those depend on the proximity to the threshold. But so long as you can estimate the ETI using the existing methods, explicit consideration of the tax-rate change is unnecessary.

Equation 14 applies for a given taxpayer, so strictly requires estimation of an individual-level treatment effect, which in practice is not observed. Instead, I must rely on an average treatment effect (ATE) among all bunchers, which introduces two imperfections: the formula is nonlinear, so the average elasticity, which I wish to measure, differs from the elasticity of the average taxpayer, which I actually measure; and I must rely on different treatment and comparison groups to construct the ATE, which could introduce a selection bias. Little can be done about the former issue, a problem that afflicts all bunching studies in which the ETI is heterogeneous, but the small range of taxable income considered suggests the bias should be modest. The latter imperfection is a major consideration in the research design, to follow.

To estimate the deduction elasticity, it is necessary to estimate the ETI, two ATEs, and two

⁹The disaggregated model I consider generates ETI heterogeneity because the ETI depends on the proportions of gross income and deductions in the tax return, which differs across taxpayers. As noted by Kleven (2016), the standard bunching model is robust to ETI heterogeneity because the ETI estimate can be interpreted as the average ETI across taxpayers. The same is true in the present case.

¹⁰Kleven and Waseem (2013) assume isoelastic and quasilinear utility in order to yield a closed-form solution for the ETI. They also propose an alternative, non-parameterised version of the formula, which relies on an inference of the approximate MTR change implicit in the average tax rate change at the threshold.

average outcomes under nontreatment:

$$\hat{e}_d = \hat{e} \cdot \frac{\hat{\mathbb{E}}[d_1 - d_0]}{\hat{\mathbb{E}}[z_1 - z_0]} \cdot \frac{\hat{\mathbb{E}}[z_0]}{\hat{\mathbb{E}}[d_0]}, \quad (15)$$

in which all expectations are conditional on $z_0 \in (z^*, z^* + \Delta z^*]$. Estimating the means requires observing taxpayers who face the notch and taxpayers who don't. In practice, discontinuities in the tax schedule often apply only to those with a particular characteristic, so that those to whom the notch applies can be compared to those to whom it does not. The Earned Income Tax Credit in the U.S., for example, applies only to those with children. Another potential source of variation is the introduction or removal of the discontinuity at a point in time. In that case, an event-study design can be used in which the tax returns of those present before the change can be compared to those present after. That is my approach.

In practice, the taxpayers who bunch do not all do so precisely. Instead, there tends to be a sharp spike in mass at the threshold, with a diffusion of excess mass throughout a region below the threshold, as shown in figure 2. The 'manipulation region', $\mathcal{Z} = [z_L, z_U]$, is the range of taxable income around the threshold containing the origin and destination of all bunchers. The upper bound of the manipulation region, z_U , corresponds to the counterfactual taxable income of the marginal buncher, $z^* + \Delta z^*$. The threshold bisects the manipulation region into lower ($\mathcal{Z}_L = [z_L, z^*]$) and upper ($\mathcal{Z}_U = (z^*, z_U]$) portions, with taxpayers moving from the upper to the lower portion when the notch is present.

For estimating the ETI, this setting offers an advantage over the usual one, in which taxpayers are observed only when the notch is in place. The task is to estimate the counterfactual taxable income of the marginal buncher, z_U , because bunching theory connects her proximity to the threshold with the ETI. Without a counterfactual density, this is difficult to determine ocularly because the missing mass is diffuse. Kleven and Waseem (2013) address this by determining ocularly the lower bound of the manipulation region, z_L , (which is easier because the excess mass deviates more sharply from the counterfactual density), and then exploiting the equality of the excess and missing mass to estimate the upper bound of the missing mass, z_U . This is problematic when there are large extensive-margin responses (as in the present case), where those above the threshold in the absence of the notch exit the sample, causing an imbalance in the excess and missing mass (balancing them is central to the standard method). I avoid these considerations by observing a counterfactual density in a period when the notch is absent. I can estimate z_U , and thus the ETI, by determining ocularly the convergence point of the actual and observed counterfactual densities, just as Kleven and Waseem do for z_L .

To estimate the ATEs, I compare taxpayers located in the manipulation region when the notch is in place to those when it is not. I consider the entire manipulation region because

it contains the origin and destination of all bunchers. A ‘local average treatment effect’, estimated in a small neighbourhood of the threshold via a regression-discontinuity design, would not identify the ATEs because of selection into bunching: those who bunch might differ from those who do not, and that difference could be related to the outcome. By considering the entire manipulation region, I avoid any selection bias due to bunching, as the outcomes of all bunchers and nonbunchers are always included in the estimates. For the average outcomes under nontreatment, I focus only on the upper portion of the manipulation region under nontreatment because that is where all bunchers originate.¹¹

In practice, however, a simple comparison of means in the manipulation region with and without the notch might not reflect the true ATE. Even isolated from any confounding effects of other policies, deductions and gross income grow over time, and likely at different rates (the consumer price inflation rate typically differs from the wage inflation rate, for example). As a result, the gross income and deductions distributions at a given level of taxable income will differ between years in the absence of treatment. Given the fixed window of taxable income that defines the treatment and comparison groups, these year-on-year changes must be controlled for.

The difference-in-differences (DiD) design is a standard way to do so. It requires the availability of a placebo group that is not exposed to the treatment in either period. If the placebo group is comparable to the treatment group, then taxpayers in the placebo group should on average exhibit the changes in gross income and deductions that those in the treatment group would have exhibited in the absence of treatment. These differences can then be subtracted from the differences observed for the treatment group to identify the ATEs. The same approach can be taken to construct counterfactual average outcomes under nontreatment for those who receive the treatment.

The placebo group can be drawn from a range of taxable income just below the manipulation region. It is important to ensure that no confounding policy affects the placebo group in the periods considered. As they are located below the threshold, taxpayers in the placebo group face no incentive to alter their tax returns in response to the treatment. And their close proximity to the manipulation region should ensure that they are comparable to those in the treatment group. The validity of this assumption can be checked by ascertaining whether there are parallel trends in outcomes between the two groups prior to treatment. If there are non-parallel pre-trends, then they can be used to predict what the counterfactual outcome would have been, with estimated average treatment effects adjusted accordingly.

¹¹Consideration of the average within the lower region would capture the mean outcomes under treatment rather than nontreatment, but it also would capture the outcomes of those located below the threshold in the absence of treatment, who have no incentive to respond to the treatment.

4 Institutional settings, and data

I study the removal of a notch in the Australian personal income tax system, which has a structure similar to those of comparable countries like the U.S..¹² Taxable income is calculated as the taxpayer's income items less her deductible expenses.^{13,14} The system differs from some others in that there are no limits on the amount of expenses that a taxpayer may deduct. A common category is work-related expenses, for which the only limits are in the types of items that may be claimed and that the expenses were incurred in the course of earning income. For example, a work uniform with an embroidered logo is deductible, but general business attire is not, and only the portion of a computer's value used for work purposes may be deducted against income based on depreciation over its life. In the baseline group of taxpayers considered later, average deductions are \$1,897, of which 4% are for charitable giving, 6% are for the cost of managing tax affairs, and 86% are for work-related expenses. The largest work-related expenses are car expenses (37% of total deductions), clothing expenses (11%), travel expenses (6%), and education expenses (5%).

The Australian Commissioner of Taxation, who is responsible for administering the Australian tax system, has claimed based on a recent increase in the number of audits that work-related expenses are widely misreported (Jordan, 2017). “[The prevalence of work-related clothing expense deduction claims] would mean that almost half of the individual taxpayer population was required to wear a uniform—suits are not uniforms—or protective clothing, or had some special requirements for things like sunglasses and hats and a variety of other things. Half the population,” the Commissioner said in a recent speech. Work-related expenses for a car and other travel are said also to have been widely misreported. The claim is not that many of the expenses were not made, but rather that they were undertaken for the purposes of consumption rather than to generate earnings.

The income tax schedule features a tax-free threshold (AU\$6,000), and increasing marginal tax rates (15, 30, 40, and 45% at \$6,000, \$34,000, \$80,000, and \$180,000, respectively). Average full-time adult total earnings are \$64,662 (Australian Bureau of Statistics, 2009), and average taxable income is \$46,462 (Australian Taxation Office, 2009). Unlike the U.S., spouses must file separately, but provide some of the details of the spouse's tax return for the purposes of means tests that depend on family income. As in the U.S., it is common for taxpayers to use a

¹²All figures are for 2009, the relevant year in which the treatment is absent for the policy that I consider. References in the paper to '2009' refer to the 2008-09 financial year, in which taxpayers lodged their tax returns between July 1 and October 31 of 2009.

¹³Income items include wage and salary income, tips, government allowances, pension income, self-employment income, fringe benefits, business income, interest, and dividends, while deductible expenses include work-related expenses (car, travel, uniform, self-education), interest and dividend expenses (any interest paid in order to generate income, which includes interest on loans to fund investments, but not for a mortgage on the primary residence), gifts and donations, business expenses, and the cost of managing one's tax affairs.

¹⁴Among all taxpayers, 65.3% claimed work-related deductions (Australian Taxation Office, 2009).

tax preparer (71.2%), though the government provides a free online system for submitting tax returns, including pre-filing of government payments and third-party reported income.¹⁵

Australia has a government-funded universal healthcare system, similar to the U.K. 'National Health System', called 'Medicare'. This system provides subsidies for general practitioner visits, and free hospital care for most treatments. Patients seeking elective procedures often are subject to a waiting period. On top of the publicly-provided system lies a voluntary private health insurance system. Private health insurance premiums are controlled and subsidised by the government. Private health insurers incur the healthcare costs that otherwise would be borne by the public system. Patients with private health insurance might be able to receive elective procedures sooner, and receive higher-quality amenities such as a private room. Around half of Australian adults are covered by private health insurance (Australian Bureau of Statistics, 2013).

In order to encourage people to take out private health insurance, in 1998 the Australian government introduced the 'Medicare Levy Surcharge' (MLS), which is an additional 1% tax (applicable to all of taxable income) on those without private health insurance coverage whose taxable income is above a threshold.^{16,17} It is a notch because tax liability increases discontinuously at the threshold.¹⁸ The threshold depends on whether a taxpayer has dependents (either a dependent spouse or children). The threshold is \$50,000 for those without dependents, and starts at \$100,000 for those with dependents, and increased incrementally for each additional child. For a single taxpayer without children and without private health insurance, an increase in taxable income from \$50,000 to \$50,001 would increase tax liability by around \$500, implying an effective-marginal tax rate of 50,000%.

I use the Australian Taxation Office unit record file, which is a 16% random sample of all taxpayers (around 4 million returns per year), including all items in the income tax return.¹⁹ The data are partly self reported and partly third-party reported. To condition on the characteristics of the taxpayers to whom the policy applies, I exclude those with a spouse (must be living with the taxpayer), dependent children, private health insurance, or eligibility for a 'Medicare levy exemption category'. I consider the threshold for those without dependents because the data are not disaggregated for the spouse.

¹⁵18.8% of tax returns were submitted by the taxpayer online (Australian Taxation Office, 2009).

¹⁶Stavrunova and Yerokhin (2014) study the effect of this policy on private health insurance take up.

¹⁷The tax does not strictly apply to taxable income, but rather to 'income for MLS purposes'. This is taxable income plus fringe benefits, 'the amount on which family trust distribution tax has been paid', and 'any element of a superannuation [similar to an IRA] lump sum for which the tax rate is zero'. For the vast majority of taxpayers, the two concepts are the same.

¹⁸It is a 'proportional notch' because both average and marginal tax rates are discontinuous at the threshold (put differently, the policy combines a 'pure notch' with a kink). The 1% increase in the marginal tax rate should cause a leftward shift in the taxable-income density above the threshold, but, as is common in studies of proportional notches, I ignore this shift in the empirical analysis later on. The small change in the marginal tax rate suggests any movement in the density will be small.

¹⁹The data are not publicly available.

The threshold for those without dependents remained at \$50,000 for 10 years. As the mass of taxpayers moved upward due to wage inflation, the tax—originally intended only for those on high incomes—came to apply to a large portion of the population. To address this concern, in 2009 the government raised the threshold from \$50,000 to \$70,000, and indexed it based on wage inflation. The policy otherwise was unaltered. The historical path of the threshold is displayed in figure 3, with the threshold change highlighted in grey. There is no change in the marginal tax rate near the original threshold of \$50,000 in the years leading up to the change, and no other policy changes apply near the threshold in the years considered.

The movement in the MLS threshold provides the necessary variation in treatment status to identify the deduction and gross-income elasticities using the bunching decomposition method. The notch is removed rather than introduced, so to minimise confusion in interpreting the results, I set the treatment period to 2009 (when the notch is absent) and the nontreatment period to 2008 (when the notch is present). The estimates should therefore be interpreted as the effects of removing the notch, which have opposite signs to the effects of introducing the notch.

I consider the \$50,000 threshold rather than the \$70,000 threshold. The policy encourages those earning above the threshold to take up private health insurance, and this incentive increases with income. In 2009, taxpayers earning \$70,000 without private health insurance pay a tax of \$700, compared to the \$500 tax paid by those earning \$50,000. The large extensive-margin response this induces for taxpayers without private health insurance earning around \$70,000 in 2008, when the tax applied from \$50,000, means they might be poor counterfactuals for those in 2009, when the threshold was raised to \$70,000. There is no such problem with those without private health insurance earning around \$50,000 in 2009, as the policy no longer applies to them in any way. Those earning \$50,000 are also closer to the peak of the taxable-income density, which increases statistical power.

The estimates apply to those who have an incentive to bunch, so it is necessary to determine who they are. With the notch, the bunchers locate below, rather than above, the notch, causing the taxable-income densities of the treated and nontreated taxpayers to diverge within a ‘manipulation region’, shown in figure 2. I describe in appendix A how I determine this region. Determining the range in which the treatment- and nontreatment-period densities diverge is equivalent to determining that in which their ratio diverges from one. A local-logit estimate of the ratio of the two densities is displayed in figure 4.²⁰ The manipulation region is located where the slope of the density ratio diverges from (approximately) zero. I determine this ocularly as the taxable incomes between \$49,150 and \$51,400, and assess the implications of this choice in a sensitivity analysis. This defines the treatment group. For the placebo

²⁰There appears to be an extensive-margin response to the right of the threshold, with the treated mass less than the nontreated mass. This likely reflects the choice of some taxpayers to take up private health insurance. I estimate the effect of any extensive-margin response to be minor, and detail my approach in appendix B.

group, I include all taxpayers with a taxable income between \$42,000 and \$47,000, which is chosen to be close to the treatment group so as to be comparable, but not so close as to be tainted by the treatment, as well as far away from any confounding policy variation further below the threshold. I denote the placebo region $\mathcal{Z}^0 = [z_L^0, z_U^0]$.

Taxpayers in 2008 are indicated by $T = 0$, those in 2009 by $T = 1$, those with taxable incomes between \$49,150 and \$51,400 by $S = 1$, and those with taxable incomes between \$42,000 and \$47,000 by $S = 0$. A visual comparison of the densities of taxable income for the placebo and treatment groups across the treatment and nontreatment periods is presented in figure 5. The region of taxable income occupied by the placebo group appears to be free of manipulation either due to the treatment or due to any confounding policies, with no difference between the treatment-period and nontreatment-period densities for the placebo group. The nontreatment-period density also appears to offer an appropriate counterfactual for the treatment-period density. Of note is a complete absence of bunching for the treatment group immediately after the removal of the notch, suggesting taxpayers are highly responsive to the policy change.

Summary statistics are displayed in table 1 for the variables of interest across the four groups. The deductions distribution has a long right tail. While median deductions are around \$1,000, some taxpayers have deductions in excess of \$200,000. The 99th percentile of deductions for all groups combined is \$16,469. To address the effect of outliers in the deductions distribution on the results, I present the results for both all taxpayers and excluding the top 1% of deduction claimers.

5 Empirical analysis

5.1 Probability of bunching

Some taxpayers near the threshold might in practice not respond to the treatment because of adjustment frictions or ETI heterogeneity (Kleven and Waseem, 2013).²¹ The bunching decomposition method relies on the standard bunching method to estimate the ETI, which is then decomposed into the item responses. This ETI estimate is for the marginal buncher, who is representative of the bunchers rather than the nonbunchers. Consistent with this, it is appropriate to condition the item responses on the bunchers, which won't affect the relative proportional changes of the items and taxable income (as both the numerator and denominator would be attenuated equally by the prevalence of nonbunchers), but might affect the estimated average outcomes under nontreatment (if the bunchers and nonbunchers differ in that dimension), and thus the deduction and gross-income elasticities.

²¹This may be likened to 'compliance', in the program evaluation parlance, in which not all units assigned to the treatment comply with their assignment.

As shown in figure 2, three groups of taxpayers are located in the manipulation region before and after treatment: those below the threshold in the absence of treatment for whom there will be no treatment effect; those above the threshold in the absence of treatment but who do not respond so for whom there will be no treatment effect; and those above the threshold in the absence of treatment but who do respond so for whom there will be a treatment effect. By default, an average in the manipulation region is for the three groups combined. But, as there is no response among two of the groups, the response among the bunchers can be determined using the proportion of taxpayers who bunch.

Bunching is indicated by $B = 1$. Given that $\mathbb{E}[d_1 - d_0 \mid B = 0] = 0$, the ATEs can be inflated so as to be conditional on bunching:

$$\mathbb{E}[d_1 - d_0 \mid z_0 \in \mathcal{Z}, B = 1] = \frac{\mathbb{E}[d_1 - d_0 \mid z_0 \in \mathcal{Z}]}{\mathbb{P}[B = 1 \mid z_0 \in \mathcal{Z}]}.$$
 (16)

For the average outcomes under nontreatment, the approach is similar. Given that $\mathbb{E}[d_1] = \mathbb{E}[d_0 \mid B = 0]$, the effect of the nonbunchers can be subtracted from the nontreated mean, which can be inflated in the same way:

$$\begin{aligned} & \mathbb{E}[d_0 \mid z_0 \in \mathcal{Z}_U, B = 1] \\ &= \frac{\mathbb{E}[d_0 \mid z_0 \in \mathcal{Z}_U] - (1 - \mathbb{P}[B = 1 \mid z_0 \in \mathcal{Z}_U]) \cdot \mathbb{E}[d_1 \mid z_0 \in \mathcal{Z}_U]}{\mathbb{P}[B = 1 \mid z_0 \in \mathcal{Z}_U]}. \end{aligned}$$
 (17)

Equations 16 and 17 require estimates of the bunching probability, which can be obtained from the densities of the treated and nontreated taxpayers. If the distribution under nontreatment is a valid counterfactual for that under treatment, then the bunching probability is:

$$\mathbb{P}[B = 1 \mid z_0 \in \mathcal{Z}] = \frac{H_0(z_U) - H_0(z^*)}{H_0(z_U) - H_0(z_L)} - \frac{H_1(z_U) - H_1(z^*)}{H_0(z_U) - H_0(z_L)},$$
 (18)

where $H_0(\cdot)$ is the cumulative distribution function of taxable income under nontreatment. This corresponds to the missing mass above the threshold in the treated density. The probability in the upper portion of the manipulation region, to be used to condition the average outcomes under nontreatment on bunching, is the same as that in equation 18, except the denominators are conditional on $z_0 \in \mathcal{Z}_U$. Equation 18 can be estimated via the corresponding empirical distributions:

$$\hat{\mathbb{P}}[B = 1 \mid T = 1, S = 1] = \frac{\sum_{i=1}^n \mathbf{1}[z_i > z^*] \cdot (1 - T_i) \cdot S_i}{\sum_{i=1}^n (1 - T_i) \cdot S_i} - \frac{\sum_{i=1}^n \mathbf{1}[z_i > z^*] \cdot T_i \cdot S_i}{\sum_{i=1}^n (1 - T_i) \cdot S_i},$$
 (19)

with standard errors computed via a bootstrap procedure.^{22,23}

If the density of taxable income in the treatment and nontreatment periods were to differ in the absence of treatment, then a DiD approach can be used in which the distribution of the treatment group in the nontreatment period is adjusted for any change over time in the placebo distribution. For identification, it must be the case that the growth rates of the distributions of the placebo and treatment groups would have been the same in the absence of treatment:

$$\begin{aligned} & \frac{H_0(z_U | T = 0, S = 1) - H_0(z_L | T = 0, S = 1)}{H_0(z_U | T = 1, S = 1) - H_0(z_L | T = 1, S = 1)} \\ &= \frac{H_0(z_U^0 | T = 0, S = 0) - H_0(z_L^0 | T = 0, S = 0)}{H_0(z_U^0 | T = 1, S = 0) - H_0(z_L^0 | T = 1, S = 0)}. \end{aligned} \quad (20)$$

The assumption for the upper portion of the manipulation region of bunching is the same, except that the distributions of the treatment group ($S = 1$) are conditional on $z_0 \in \mathcal{Z}_U$. The corrected bunching probability can then be estimated by multiplying the second fraction in equation 19 by:

$$\frac{\hat{H}(z_U^0 | T = 0, S = 0) - \hat{H}(z_L^0 | T = 0, S = 0)}{\hat{H}(z_U^0 | T = 1, S = 0) - \hat{H}(z_L^0 | T = 1, S = 0)} = \frac{\sum_{i=1}^n (1 - T_i) \cdot (1 - S_i)}{\sum_{i=1}^n T_i \cdot (1 - S_i)}.$$

The bunching probability estimates are presented in table 2. The ‘DiD’ columns are corrected for the observed change in the placebo density.

5.2 Changes in the items when bunching

In the DiD design, identification of the ATEs requires the difference in nontreated outcomes between the treatment and placebo groups to be constant over time:

$$\begin{aligned} & \mathbb{E}[d_0 | T = 1, S = 1] - \mathbb{E}[d_0 | T = 1, S = 0] \\ &= \mathbb{E}[d_0 | T = 0, S = 1] - \mathbb{E}[d_0 | T = 0, S = 0]. \end{aligned} \quad (21)$$

A potential threat to identification is non-parallel pre-trends in the average outcomes of the treatment and comparison groups. In the context I consider, data are available only for the three years prior to treatment. Observe in figure 6 that, for taxable income, the within-year difference between the treatment and placebo groups appears to be constant prior to

²²In the bootstrap procedure, I draw with replacement 10,000 random samples (each with a sample size equal to the full sample size), then perform the estimation procedure in equation 19 on each of the samples. The reported standard errors are the standard deviations of the estimated probabilities across the 10,000 samples.

²³To obtain the probability of bunching among taxpayers in the upper portion of the manipulation region, which is used to condition the levels of the outcomes under nontreatment among bunchers, the same procedure may be used but with an indicator function included in the denominators given by $\mathbf{1}[z_i > z^*]$.

treatment, consistent with parallel pre-trends. This is unsurprising, as the same band of taxable income is chosen in each year. The differences for gross income and deductions, however, appear to narrow over time.

As the placebo group is located in a lower region of the taxable income distribution, those taxpayers also have on average lower deductions and gross income, both of which are increasing functions of taxable income. Even if the inflation rates of the items are independent of taxable income, they are being applied to a lower base for the placebo group. This means that the year-on-year changes in gross income and deductions would be higher for the placebo group than for the treatment group, leading to a convergence of the two over time. The secular nature of this process suggests it should be straightforward to address, aided by the apparent linearity of the time trend.

I estimate the ATEs under three specifications. First, I compute a simple difference in average outcomes in the manipulation region between the treatment and nontreatment periods. As there is reason to doubt the exogeneity of the treatment, I also compute standard DiD estimates. Then, to address non-parallel pre-trends, I use a linear regression of the outcomes in the three years prior to the treatment period to predict the nontreatment-period outcomes, which I substitute for the observed outcomes when calculating the ATEs. The latter generates my preferred estimates.

The three specifications rely on two regression models. The first is a standard DiD regression of the form:

$$d_i = \beta_0 + \beta_1 \cdot T_i + \beta_2 \cdot S_i + \beta_3 \cdot (T_i \times S_i) + \varepsilon_i. \quad (22)$$

For the first specification, the estimate of the ATE is $\hat{\beta}_1$, and, for the second specification, it is $\hat{\beta}_3$. For the third specification, let the year an observation is observed, among the three prior to treatment, be $Y_i \in \{2006, 2007, 2008\}$, and then let:

$$y_i = \begin{cases} -1 & \text{if } Y_i = 2008 \\ -2 & \text{if } Y_i = 2007 \\ -3 & \text{if } Y_i = 2006. \end{cases}$$

The second regression is:

$$d_i = \delta_0 + \delta_1 \cdot S_i + \delta_2 \cdot y_i + \delta_3 \cdot (y_i \times S_i) + \gamma_i. \quad (23)$$

For the third specification, the estimate of the ATE is then $\hat{\beta}_2 + \hat{\beta}_3 - \hat{\delta}_1$.

Regression output for the three specifications is presented in table 3, and the estimates for the DiD specifications are depicted in figure 6. The columns titled 'Difference', 'DiD (1)', and

'DiD (2)' contain estimates under the first, second, and third specifications, respectively. To address outliers in the deductions distribution, I compute the first two both on all taxpayers, and on a group excluding the top 1% of deduction claimers, and the third specification only on the latter. In the final column of the table are estimates computed under the third specification but conditional on the bunchers. For these, the standard errors are estimated via a bootstrap procedure.²⁴ I find the effect on the results of any extensive-margin response to be modest, and describe my approach in appendix B.

The DiD estimates are similar across the trimmed and full samples, but the standard errors for the former are substantially lower. The simple differences appear to misrepresent substantially the treatment effect. As can be seen in figure 6, there is some variation in outcomes across years, but the variation of the treatment and placebo groups is highly correlated, which supports the DiD design. The presence of converging pre-trends appears to bias the estimates, overstating the role of deductions and understating that of gross income.

The differences-in-differences of the three outcomes, with deductions represented in terms of their contribution to taxable income (that is, negative deductions), are depicted in figure 7. Under parallel pre-trends, the differences in differences in the years prior to treatment would be zero. As shown in figure 7a, this is not the case for gross income and deductions, but controlling for linear pre-trends shifts the plots for gross income and deductions vertically so that they are centered at zero, as shown in figure 7b. Deductions account for around a third of the response in taxable income, with gross income accounting for the remaining two thirds.

5.3 Levels of the items before bunching

For the average outcomes under nontreatment, the identification assumptions and regression specifications are the same as those for the ATEs. The analysis focuses, among the treatment group, only on those taxpayers with a taxable income above the threshold. The counterfactuals I construct are estimates of what the outcomes would have been in the treatment period in the absence of treatment.

For the first specification, the estimate is $\hat{\beta}_0 + \hat{\beta}_1$ from the model in equation 22, but, for those in the treatment group, only among those above the threshold. For the second specification, the estimate is $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2$ from the same model. For the third specification, the estimate is $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 - \hat{\delta}_3$ from the models in equations 22 and 23. In this specification, the difference in slopes between the treatment and placebo groups over the three-year period prior to treatment is deducted from the treatment effect to account for pre-trends.

²⁴In the bootstrap procedure, I draw with replacement 10,000 random samples (each with a sample size equal to the full sample size), then perform the estimation procedure in equation 19 on each of the samples. I then divide the ATE estimate by the bunching probability estimate. The reported standard errors are the standard deviations of the ATE conditional on bunching across the 10,000 samples.

Regression output for the three specifications is presented in table 4, and construction of the counterfactuals is depicted in figure 8. The columns titled ‘Actual’, ‘C.f. (1)’, and ‘C.f. (2)’ contain estimates under the first, second, and third specifications, respectively. In the final column estimates are computed under the third specification, but conditional on the bunchers. For these, the standard errors are estimated via a bootstrap procedure.

There are large differences in outcomes under nontreatment between the bunchers and nonbunchers, which is indicative of selection into bunching. Bunchers are estimated to have 12% higher deductions in the absence of treatment, for example. This highlights the importance of focusing on changes in average outcomes within the entire manipulation region, rather than locally at the threshold or above the threshold, where any estimated treatment effects would be biased by selection into bunching. The counterfactuals are displayed in figure 8. As one would expect, the first counterfactual remains parallel to the levels in the placebo group between 2006 and 2009. This appears to be inappropriate, as the levels of both gross income and deductions converge prior to treatment. The second counterfactual, which adds the convergence in trends back to the counterfactual, appears to perform better.

5.4 Deduction and gross-income elasticities

The key results of the paper are presented in table 5.²⁵ The sufficient statistics of interest are the deduction and gross-income elasticities with respect to the net-of-tax rate, which consist of the elasticities with respect to taxable income, driven by the relative proportional changes of deductions and gross income, and taxable income, multiplied by the ETI. The first column displays the proportion of deductions and gross income in taxable income in the absence of treatment, and the second the proportion of the change in taxable income that can be attributed to them. Dividing the second column by the first yields the item elasticity with respect to taxable income. Then multiplying by the ETI yields the deduction and gross-income elasticity with respect to the net-of-tax rate.

Deductions constitute 5% of taxable income, but account for 35% of the response of taxable income to the tax rate. Accordingly, a 1% decrease in taxable income is achieved via a 7.5% increase in deductions, and only a 0.62% decrease in gross income. The ETI is a scalar, having no effect on the relative item elasticities. Using the reduced-form approximation of Kleven and Waseem (2013), the upper bound of the manipulation region (\$51,400) implies an

²⁵These results are computed using: the trimmed sample, which excludes the top 1% of deduction claimers; the ‘DiD’ specification for the bunching probability estimation, which accounts for the change over time in the placebo density; the pretrend-corrected estimates for both the changes and levels of the outcome; and the estimates conditioned on bunching, which differ only due to the different levels in the absence of treatment between the bunchers and nonbunchers.

ETI of 0.06.²⁶ This yields a gross-income elasticity of 0.04, and a deduction elasticity of -0.45 .

While deductions account for only a small fraction of taxable income, they account for a large fraction of the response of taxable income to the tax rate, making the deduction elasticity more than an order of magnitude larger than the gross-income elasticity. This suggests that a reduction in deductibility could substantially reduce the ETI, and thus the impact of income taxation on welfare. Whether the ETI would fall all the way to equal the gross-income elasticity depends on the validity of the separability assumption.

Using the formula for the optimal deductibility rate (equation 9), it is possible to determine the external benefit that deductions would need to generate in order for full deductibility to be optimal. Assuming that the marginal tax rate is 0.315, as it is where the notch is located, and that the marginal social utility cost of public funds (that is, the opportunity cost of \$1 of funds held by government) is 1.2 (implying a marginal efficiency cost of revenue collection of 20%), then the marginal dollar of deductions would need to generate at least 68¢ in external value over and above the benefit to the taxpayer claiming them.²⁷ If \$1 of deductions were to generate, for example, 30¢ in external benefits, the optimal deductibility rate would be just 34%.

6 Conclusion

The composition of the response of taxable income to the tax rate, not just its magnitude, is informative for policy. While many scholars have studied the ETI, and generally found it to be modest, there has not been an appreciation of the sources of the response. My results suggest that, if gross income and deductions are separable, or close to it, then a substantial lowering of the deductibility rate would substantially lower the ETI. This is similar to the argument of Slemrod and Kopczuk (2002), who show that the ETI is endogenous to the size of the tax base. That such a large proportion of the ETI is driven by something as small as deductions suggests a policy change is in order. Indeed, it seems inconceivable that the same marginal tax rate should apply to two items for which the behavioural responses are more than an order of magnitude apart.

My results of course apply to the Australian tax system. The observed elasticities are a function of the setting—the tax law, enforcement regime, local tax morale, among countless other things. The relatively liberal conditions governing the deductions that can be claimed

²⁶From Kleven and Waseem (2013), the reduced-form approximation of the ETI is:

$$e \approx \frac{((z_u - z^*)/z^*)^2}{\Delta t / (1 - t)} = \frac{(\$1,400/\$50,000)^2}{0.01/0.685} = 0.057.$$

²⁷Even if the efficiency cost of revenue collection were only 10%, deductions would still need to generate 48¢ in external value in order for full deductibility to be optimal.

are sure to play a role, and the results will differ in other countries, such as the U.S., where the rules are tighter. That a fairly large deduction elasticity has been estimated too in Germany and Austria lends some weight to the external validity of the findings. But I offer a set of tools that can be applied wherever a notch is introduced or removed, supporting replicability in other settings.

The recent focus on the ETI has been motivated by Feldstein's (1999) argument that it summarises the effect of the income tax rate on welfare. In reality, however, the government has more tax instruments at its disposal than just the income tax rate. It is within the government's remit to decide the extent to which taxpayers may claim deductions. But it could just as easily choose the extent to which the tax rate applies to all of the other items in the tax return. For this richer set of choices, the ETI is not sufficient. In this paper, I propose a method for decomposing the ETI into the item elasticities. Armed with estimates of more of the item elasticities, governments could make changes to the effective item-specific tax rates that raise welfare.

References

- Australian Bureau of Statistics (2009), Average weekly earnings, Technical Report 6302.0, Australian Bureau of Statistics. Available here: [http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/376AFC773288BA43CA25761000197A9F/\\$File/63020_may%202009.pdf](http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/376AFC773288BA43CA25761000197A9F/$File/63020_may%202009.pdf).
- Australian Bureau of Statistics (2013), Australian health survey: Health service usage and health related actions, 2011-12, Technical Report 4364.0.55.002, Australian Bureau of Statistics. Available here: <http://www.abs.gov.au/ausstats/abs@.nsf/lookup/E334D0A98272E4DCCA257B39000F2DCF?opendocument>.
- Australian Taxation Office (2009), Taxation statistics 2008-09, Technical report, Australian Taxation Office. Available here: <https://www.ato.gov.au/assets/0/104/300/362/2cb0b6c4-67b5-47a5-b4c1-b1554a05e7d1.pdf>.
- Bachas, P. and Soto, M. (2017), 'Not(ch) your average tax system: Corporate taxation under weak enforcement'. Unpublished working paper, available here: https://dl.dropboxusercontent.com/content_link/siuYy8luN6NcAbjydJrg0HLEy5E12TGeXKrI1fBap92PB6t0rffhyM0qNSd2K9E9/file.
- Best, M., Brockmeyer, A., Kleven, H., Spinnewijn, J. and Waseem, M. (2015), 'Production versus revenue efficiency with limited tax capacity: Theory and evidence from Pakistan', *Journal of Political Economy* **123**(6), 1311–1355.
- Bovenberg, A. L. and de Mooij, R. A. (1994), 'Environmental levies and distortionary taxation', *American Economic Review* **84**(4), 1085–1089.

- Chetty, R. (2009), 'Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance', *American Economic Journal: Economic Policy* **1**(2), 31–52.
- Chetty, R., Friedman, J. N., Olsen, T. and Pistaferri, L. (2011), 'Adjustment costs, firm responses, and micro vs. macro labor supply responses: Evidence from Danish tax records', *The Quarterly Journal of Economics* **126**(2), 749–804.
- Doerrenberg, P., Peichl, A. and Siegloch, S. (2015), 'The elasticity of taxable income in the presence of deduction possibilities', *Journal of Public Economics* **151**, 41–55.
- Feldstein, M. (1995), 'The effect of marginal tax rates on taxable income: A panel study of the 1986 Tax Reform Act', *Journal of Political Economy* **103**(3), 551–572.
- Feldstein, M. (1999), 'Tax avoidance and the deadweight loss of the income tax', *The Review of Economics and Statistics* **81**(4), 674–680.
- Frölich, M. (2006), 'Non-parametric regression for binary dependent variables', *The Econometrics Journal* **9**(3), 511–540.
- Gruber, J. and Saez, E. (2002), 'The elasticity of taxable income: Evidence and implications', *Journal of Public Economics* **84**, 1–32.
- Jordan, C. (2017), 'Commissioner's address to the National Press Club'. Transcript available here: <https://www.ato.gov.au/Media-centre/Speeches/Commissioner/Commissioner-s-address-to-the-National-Press-Club/>.
- Kleven, H. (2016), 'Bunching', *Annual Review of Economics* **8**, 435–464.
- Kleven, H. and Waseem, M. (2013), 'Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan', *The Quarterly Journal of Economics* **128**(2), 669–723.
- Paetzold, J. (2017), 'How do wage earners respond to a large kink? Evidence on earnings and deduction behavior from Austria'. Unpublished working paper, available here: https://www.uni-salzburg.at/fileadmin/multimedia/SOWI/documents/VWL/Working_Papers/Workingpaper_2017_01.pdf.
- Ramsey, F. (1927), 'A contribution to the theory of taxation', *The Economic Journal* **37**(145), 47–61.
- Rosenbaum, P. R. and Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.
- Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal of Educational Psychology* **66**(5), 688–701.
- Saez, E. (2010), 'Do taxpayers bunch at kink points?', *American Economic Journal: Economic Policy* **2**(3), 180–212.

- Saez, E., Slemrod, J. and Giertz, S. H. (2012), 'The elasticity of taxable income with respect to marginal tax rates: A critical review', *Journal of Economic Literature* **50**(1), 3–50.
- Schächtele, S. (2016), 'Deduction responses to the income tax: Bunching evidence from Germany'. Unpublished working paper, available here:
https://www.econstor.eu/bitstream/10419/145748/1/VfS_2016_pid_6770.pdf.
- Slemrod, J. (1998), 'Methodological issues in measuring and interpreting taxable income elasticities.', *National Tax Journal* **51**(4), 773–788.
- Slemrod, J. and Kopczuk, W. (2002), 'The optimal elasticity of taxable income', *Journal of Public Economics* **84**(1), 91–112.
- Stavrunova, O. and Yerokhin, O. (2014), 'Tax incentives and the demand for private health insurance', *Journal of Health Economics* **34**, 121–130.
- Weber, C. (2014), 'Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences', *Journal of Public Economics* **117**, 90–103.

Appendices

A Nonparametric density ratio estimation

Determining the range of taxable income in which the densities in the treatment and non-treatment periods diverge is equivalent to determining that in which their ratio diverges from one. With that in mind, note that Bayes' rule implies that the ratio of the two densities may be estimated as:

$$\frac{\hat{h}(z | T = 1)}{\hat{h}(z | T = 0)} = \frac{\hat{\mathbb{P}}[T = 1 | z]}{1 - \hat{\mathbb{P}}[T = 1 | z]} \cdot \frac{1 - \hat{\mathbb{P}}[T = 1]}{\hat{\mathbb{P}}[T = 1]}. \quad (24)$$

I estimate these conditional probabilities via local-likelihood logit regression, which is a standard nonparametric local regression but with a logistic rather than simple linear specification for the local regressions, and estimation via maximum likelihood (Frölich, 2006). Local regression has the attractive property of performing well at endpoints, such as at the threshold in the present context, hence its popular use in regression-discontinuity designs. For each value of $z = \tilde{z}$, the local-likelihood logit estimator is given by $\hat{\mathbb{P}}[T = 1 | \tilde{z}] = 1/(1 + \exp(-g(\tilde{z}; \hat{\beta}_{\tilde{z}})))$, where:

$$\hat{\beta}_{\tilde{z}} = \arg \max_{\beta_{\tilde{z}}} \sum_{i=1}^n \left(T_i \ln \left(\frac{1}{1 + e^{-g(z_i; \beta_{\tilde{z}})}} \right) + (1 - T_i) \ln \left(\frac{1}{1 + e^{g(z_i; \beta_{\tilde{z}})}} \right) \right) \times K(z_i - \tilde{z}),$$

where T_i is the treatment-period indicator for observation i , and:

$$g(z_i; \beta) = \beta_0 + \beta_1 \cdot \mathbf{1}[z_i > z^*] + \beta_2 \cdot z_i + \beta_3 \cdot (z_i \times \mathbf{1}[z_i > z^*]),$$

which, given the discontinuity at the threshold, includes taxable income, a dummy variable indicating taxpayers located above the threshold, and their interaction, and:

$$K(z_i - \tilde{z}) = \frac{3}{4} \cdot \frac{1 - \left(\frac{z_i - \tilde{z}}{h}\right)^2}{h},$$

which is the Epanechnikov kernel with a bandwidth of h . I set $h = 500$, which visually appears to provide an acceptable balance between over- and under-smoothing. This maximisation problem can be solved via maximum likelihood. The predicted values from this regression are consistent estimates of the conditional probabilities in equation 24. The unconditional treatment probabilities in equation 24 can be estimated via simple averages of the treatment indicator in the relevant ranges of taxable income considered. One then can use

these predicted values and means to compute the density-ratio estimate in equation 24.

B Extensive-margin response

The tax applies only to those who do not have private health insurance. There are two ways taxpayers can avoid the tax: by forgoing private health insurance, and reducing taxable income to below the threshold (an intensive-margin response), or by forgoing reducing taxable income, and taking up private health insurance (an extensive-margin response). If private health insurance were randomly assigned, then any extensive-margin response could be ignored for the purposes of estimating the effect of the notch on deductions and gross income. As the take up of private health insurance is a choice, it's possible that the propensity to do so is related to deductions and gross income, which would introduce a selection bias.

I observe the complement of the set of taxpayers considered for the main analysis, which includes the extensive-margin responders (the exempt group). One complication, however, is that there is mild, but visually evident bunching below the threshold among this group in the treatment period (which disappears in the nontreatment period). This must be because not all either are legally exempt from the tax, or believe that they are. One possible reason is that I exclude all those with a spouse, but not all spouses can be classified as a dependent for the purposes of the policy. Some taxpayers with a spouse might therefore face an incentive to move below the threshold to avoid the tax. This makes it difficult to distinguish between changes due to bunching, and changes due to the extensive-margin response.

I implement the following strategy to address this problem. I assume that bunchers in the exempt group have on average the same responses as the bunchers in the non-exempt group. I measure the proportion of bunchers in the manipulation region (as defined for the non-exempt group) in the exempt group (using the methods described earlier), and then use this proportion to impute the expected effect of bunching among all taxpayers (bunchers and nonbunchers) in the exempt group. I then subtract this from the observed differences-in-differences for the exempt group to determine the effect of the extensive-margin responders on deductions, gross income, and taxable income among the exempt group.

The results are displayed in table 6. The estimated probability of bunching is 5.17%. In the first column are estimates of the expected effect of bunching in the exempt group if the bunchers in the exempt and non-exempt groups had the same response (the probability of bunching multiplied by the estimates among the bunchers in the final column of table 3). The second column is the observed effect for the exempt group, based on the differences-in-differences. By subtracting the implied bunching effect from the observed effect, I obtain the implied extensive-margin effect in the final column. This captures the estimated effect of the extensive-margin responders on average deductions, gross income, and taxable income. The estimated effects are both substantively modest and statistically insignificant.

Figures

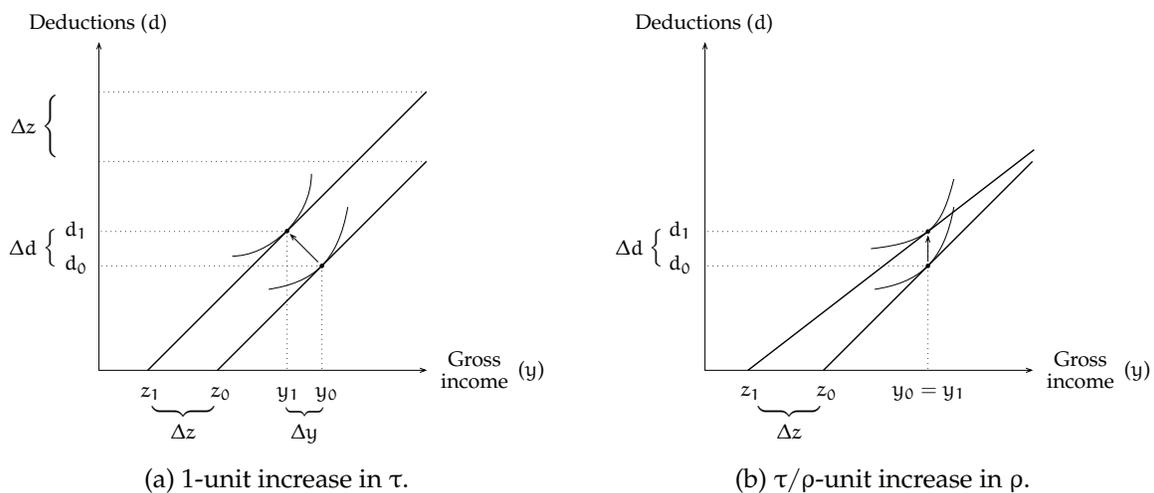


Figure 1: *Increase in the tax versus deductibility rate.* Taxable income is $z = y - d$. The responses assume quasilinear, isoelastic, and separable utility, so the the effect of the tax and deductibility rates on deductions is the same, but only the tax rate affects gross income.

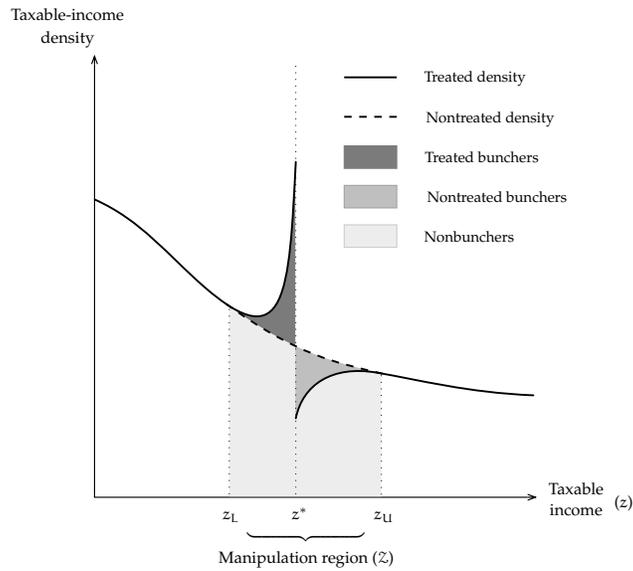


Figure 2: *Effect of manipulation on taxable-income density.* The ‘bunchers’ relocate below the threshold when treated, while the ‘nonbunchers’ don’t alter their taxable income when treated.

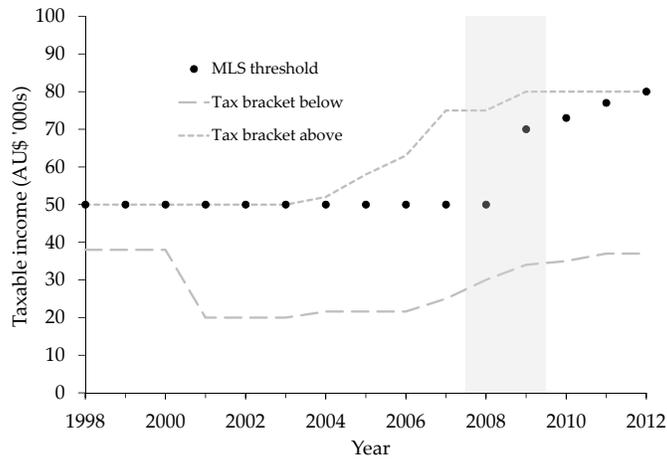


Figure 3: *Medicare Levy Surcharge threshold over time.* The surcharge was introduced in 1998, with the threshold constant at \$50k until 2009, when it was raised to \$70k, then indexed.

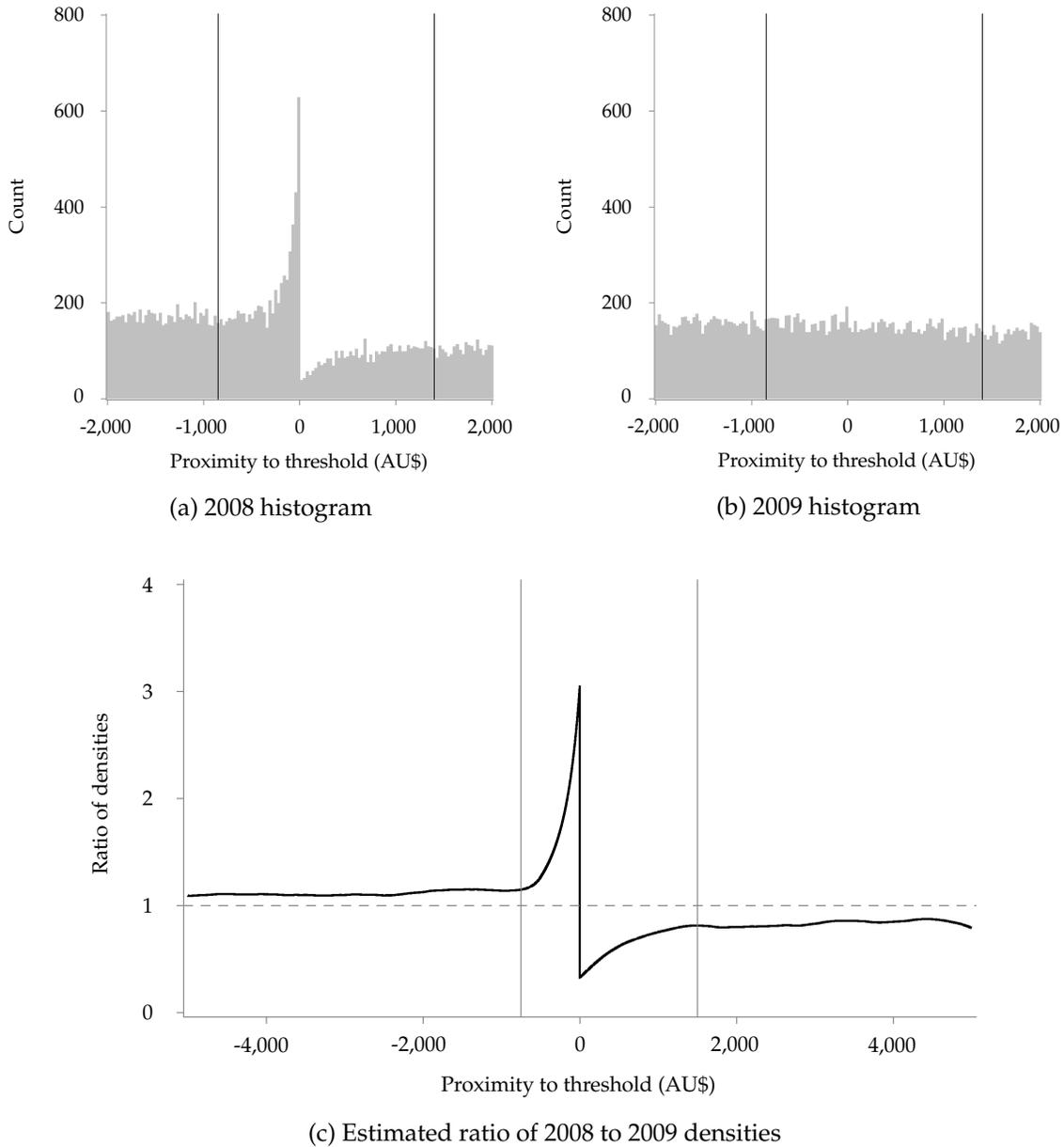
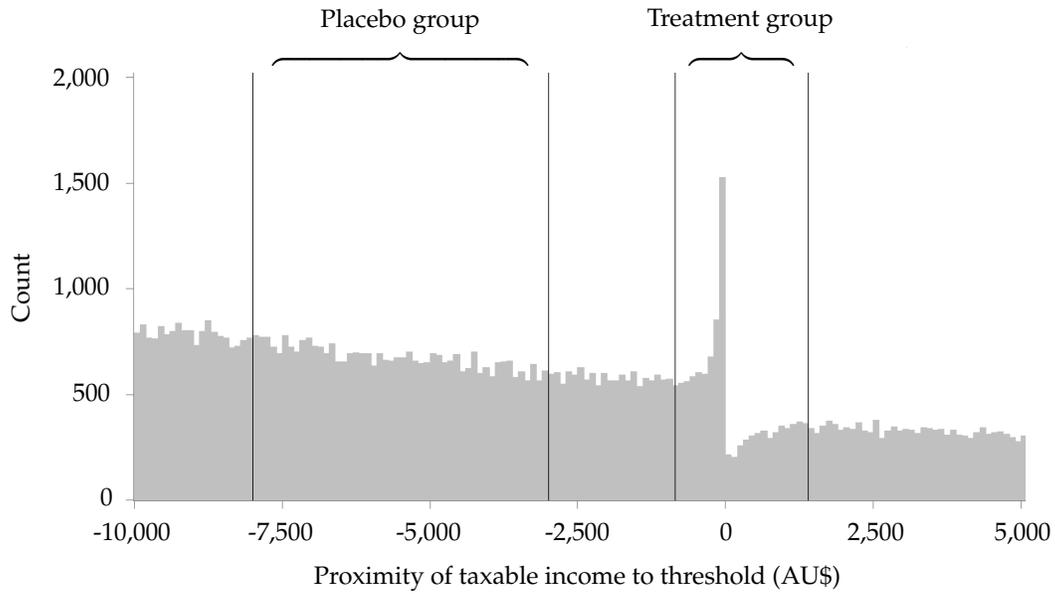
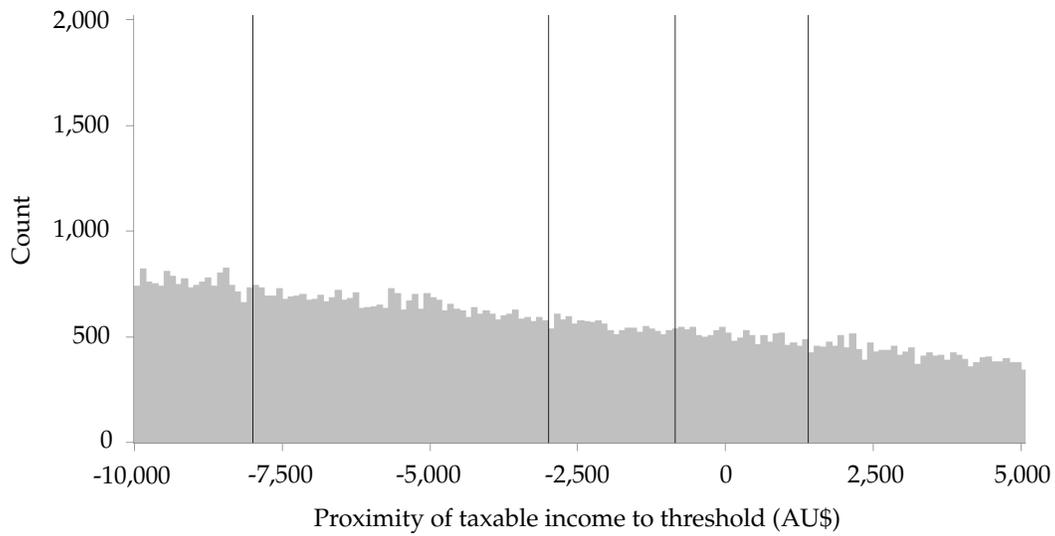


Figure 4: *Determining the manipulation region.* The top two subfigures show the histograms of taxable income under treatment and nontreatment. As the missing mass above the threshold is diffuse, it is difficult to determine ocularly the convergence point of the two densities when observing only the treated density. This is not a problem for me because I observe the nontreated density. In the present case, this is eased by observation of the nontreatment density. The bottom subfigure plots an estimate of the ratio of the treated density in subfigure 4a and the nontreated density in subfigure 4b. The vertical axis is the multiple of taxpayers in the treatment group relative to the nontreatment group. The convergence and divergence points are readily identifiable. The details of the estimation procedure are in appendix A. The vertical bars indicate the chosen manipulation region, [-850, 1,400].

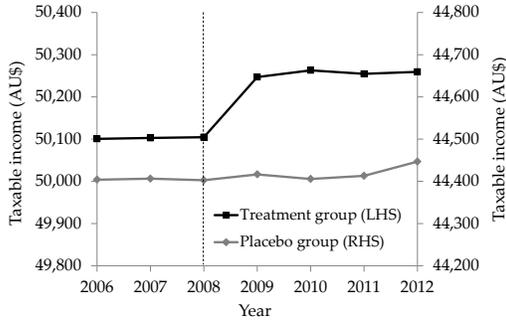


(a) Treatment period (2008).

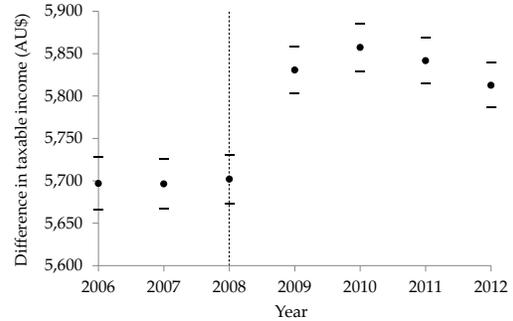


(b) Nontreatment period (2009).

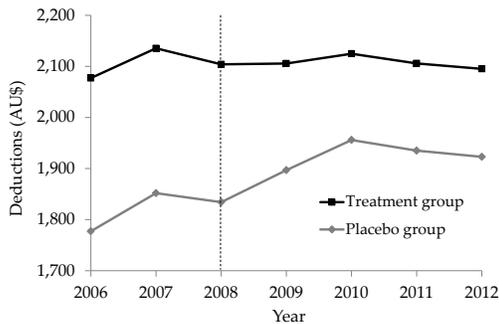
Figure 5: *Histograms of the proximity of taxable income to the threshold.* The histograms cover taxable incomes from \$40,000 to \$55,000, with a bin size of \$100. The four groups displayed are the placebo and treatment groups in the treatment and nontreatment periods. The removal of the treatment immediately eliminates the distortion present under treatment. The treatment appears to have no effect on the placebo-group density in either period.



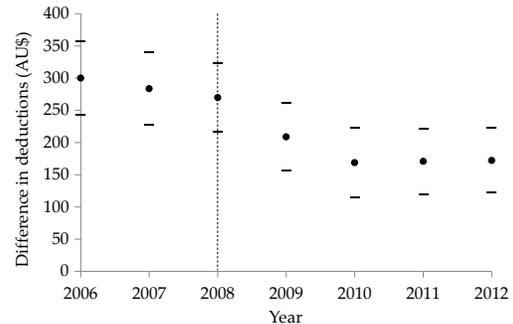
(a) Levels of taxable income



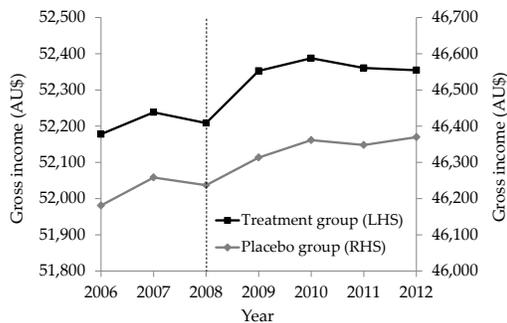
(b) Difference in taxable income



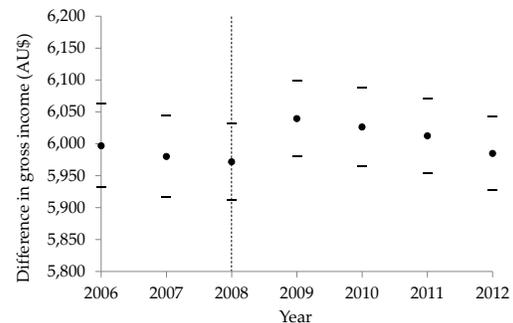
(c) Levels of deductions



(d) Difference in deductions

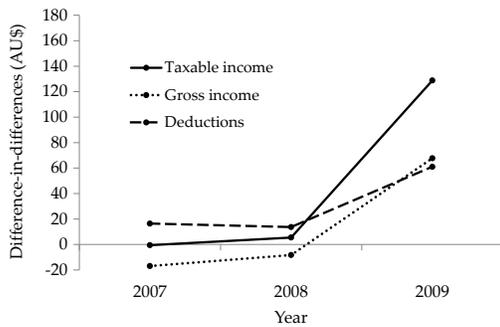


(e) Levels of gross income

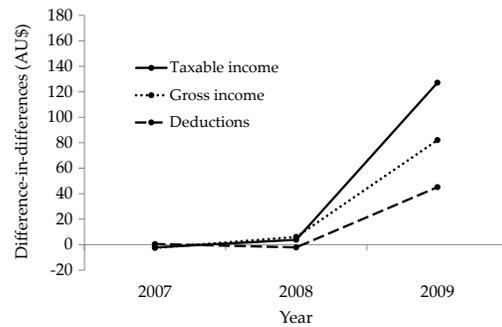


(f) Difference in gross income

Figure 6: *Checking for parallel pre-trends.* The notch is removed in 2009. Presented in figures 6a, 6c, and 6e are the average levels of the outcomes from 2006 to 2012 in the treatment and placebo groups. In figures 6b, 6d, and 6f are the differences in the average levels of the outcomes between the treatment and placebo groups, with 95% confidence interval bands. The figures exclude the top 1% of deduction claimers.

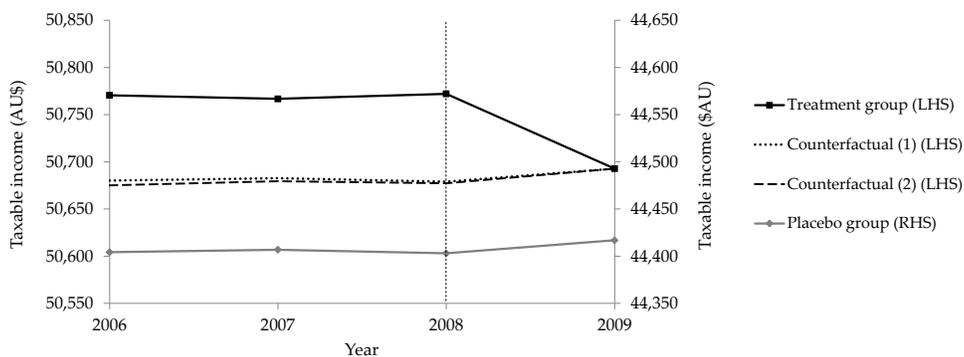


(a) Observed (DiD (1))

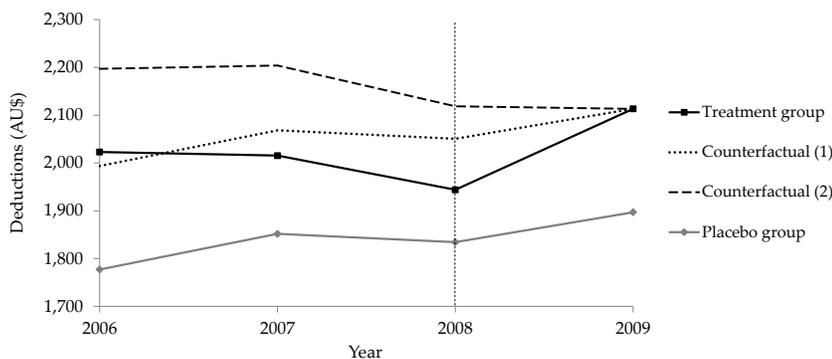


(b) Linear-pretrend corrected (DiD (2))

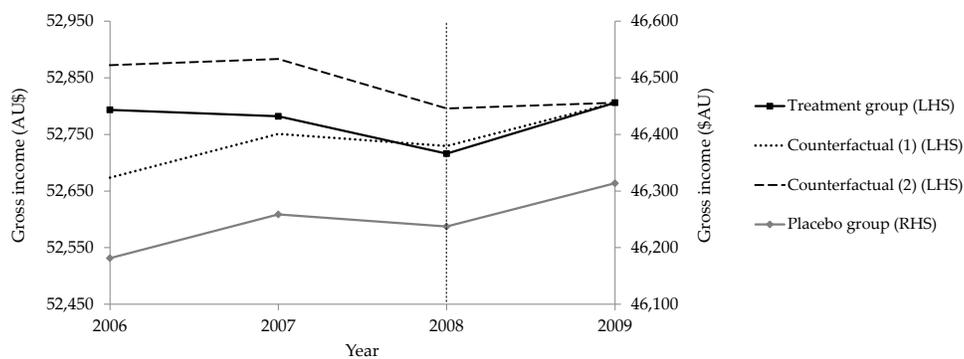
Figure 7: *Differences-in-differences over time*. The figures display the year-on-year differences-in-differences for the two items and taxable income for two years with the treatment and one year without. The negative of deductions is displayed to reflect their contribution to taxable income. In figure 7a, the displayed difference is between the observed year-on-year differences between the averages for the two groups. In figure 7b, the year-on-year differences have been corrected for linear pre-trends, which induces a parallel upward or downward shift in each of the lines compared to those displayed in figure 7a.



(a) Taxable income



(b) Deductions



(c) Gross income

Figure 8: *Counterfactual outcomes.* The grey and black solid lines indicate the observed outcomes in the placebo and treatment groups. The dotted line is the 2008 level estimated using a standard DiD, assuming the change in levels between 2008 and 2009 would have been the same in the treatment and placebo groups in the absence of treatment. The dashed line is the same as the dotted line, but corrected for linear pre-trends. The figures exclude the top 1% of deduction claimers.

Tables

	Median	p90	p99	Max.	Mean	St. dev.	N
Deductions							
Placebo, nontreatment	872	4,937	15,182	230,400	2,028	3,559	33,983
Placebo, treatment	900	5,107	14,959	137,153	2,072	3,475	33,019
Treatment, nontreatment	1,118	5,507	17,537	105,477	2,370	3,865	10,615
Treatment, treatment	1,058	5,687	16,725	110,041	2,348	3,788	11,486
Gross income							
Placebo, nontreatment	45,867	49,857	59,818	272,522	46,432	3,856	33,983
Placebo, treatment	45,902	50,088	59,827	179,511	46,490	3,780	33,019
Treatment, nontreatment	51,455	55,736	67,685	155,993	52,475	3,913	10,615
Treatment, treatment	51,537	56,096	67,317	160,807	52,595	3,841	11,486
Taxable income							
Placebo, nontreatment	44,370	46,434	46,946	47,000	44,404	1,443	33,983
Placebo, treatment	44,393	46,442	46,946	47,000	44,417	1,435	33,019
Treatment, nontreatment	49,947	51,109	51,370	51,400	50,105	632	10,615
Treatment, treatment	50,237	51,159	51,376	51,400	50,247	650	11,486

Table 1: *Summary statistics.* The four groups are: the placebo group (taxable incomes between \$42,000 and \$47,000) in the nontreatment period (2009); the placebo group in the treatment period (2008); the treatment group (taxable incomes between \$49,150 and \$52,250) in the nontreatment period; and the treatment group in the treatment period.

	Manipulation region		Above threshold	
	Difference	DiD	Difference	DiD
Bunching probability	0.2311 (0.0080)	0.2413 (0.0083)	0.3836 (0.0119)	0.4005 (0.0125)
N	15,596	82,082	11,071	77,549

Table 2: *Estimated probabilities of bunching.* ‘Difference’ is the ratio of missing mass above the threshold in the treatment period to the total mass in the manipulation region in the nontreatment period. ‘DiD’ controls for the difference in mass for the placebo group between the treatment and nontreatment periods.

	Bunchers and nonbunchers					Bunchers
	Full sample		Top 1% deductions trimmed			
	Diff.	DiD (1)	Diff.	DiD (1)	DiD (2)	DiD (2)
Deductions	-22.03 (51.50)	-66.62 (55.83)	1.58 (34.43)	-61.03 (37.97)	-45.05 (49.74)	-186.70 (215.55)
Gross income	120.44 (52.18)	62.79 (59.51)	144.16 (35.45)	67.73 (43.25)	82.05 (56.74)	340.03 (230.60)
Taxable income	142.47 (8.63)	129.40 (20.00)	142.58 (8.69)	128.76 (20.11)	127.10 (26.61)	526.73 (79.75)
N	22,101	89,103	21,860	88,338	166,111	166,111

Table 3: *Estimated average treatment effects on the treated.* ‘Bunchers and nonbunchers’ refers to estimates among all taxpayers in the manipulation region, and ‘Bunchers’ among only those who respond to the treatment. ‘Diff.’ is the average difference for the treatment group between 2008 and 2009. ‘DiD (1)’ is the difference between the treatment and placebo groups’ average differences between 2008 and 2009. ‘DiD (2)’ is corrected for linear pre-trends between 2006 and 2008. In the final column, the second-last column is divided by the probability estimate in the second of four columns in table 2.

	Bunchers and nonbunchers					Bunchers
	Full sample		Top 1% deductions trimmed			
	Actual	C.f. (1)	Actual	C.f. (1)	C.f. (2)	C.f. (2)
Deductions	2,353.25 (46.52)	2,308.66 (97.07)	2,113.27 (30.86)	2,050.65 (65.60)	2,118.53 (71.65)	2,379.74 (154.06)
Gross income	53,045.85 (46.66)	52,988.20 (104.02)	52,806.04 (31.16)	52,729.61 (75.64)	52,795.82 (82.74)	52,915.09 (160.42)
Taxable income	50,692.60 (4.84)	50,679.53 (36.25)	50,692.78 (4.86)	50,678.95 (36.45)	50,677.29 (40.07)	50,535.35 (44.32)
N	6,917	73,919	6,849	73,327	178,211	178,211

Table 4: *Estimated average outcomes among the treated under nontreatment.* ‘Bunchers and nonbunchers’ refers to estimates among all taxpayers in the manipulation region, and ‘Bunchers’ among only those who respond to the treatment. ‘Actual’ is the average level for the treatment group in 2009. ‘C.f. (1)’ is adjusted by the change in outcomes observed for the placebo group between 2008 and 2009. ‘C.f. (2)’ is corrected for linear pre-trends between 2006 and 2008. In the final column, the second-last column is divided by the probability estimate in the final column of table 2.

	% of TI	% of Δ TI	Item elasticity w.r.t.	
			Taxable income	Net-of-tax rate
Deductions	4.71	35.45	-7.53	-0.45
Gross income	104.71	64.55	0.62	0.04

Table 5: *Estimated deduction and gross-income elasticities.* The first column contains the estimated percentage of taxable income accounted for by deductions and gross income in the treatment period in the absence of treatment. The second column contains the estimated average treatment effects on the treated as a percentage of the estimated change in taxable income. The estimates in the third column are calculated by dividing the second column by the first. This yields the elasticities of deductions and gross income with respect to taxable income, which is their percentage change given a 1% change in taxable income. The estimates in the final column are calculated by multiplying the third column by the ETI estimate of 0.06. These results exclude the top 1% of deduction claimers, using the 'DiD' specification for the bunching probability, the pretrend-corrected estimates for both the changes and levels of the outcome, and conditional on the bunchers (which differ only due to the different levels in the absence of treatment between the bunchers and nonbunchers).

	Implied bunching effect	Estimated differences in differences	Implied extensive margin effect
Deductions	-9.68	-20.17 (19.36)	-10.49
Gross income	17.64	5.718 (21.90)	-11.92
Taxable income	25.89	27.32 (9.92)	-1.43

Table 6: *Estimated extensive-margin effect.* In the first column, the final column of table 3 (the estimated ATEs among bunchers) is multiplied by the estimated probability of bunching in the exempt group, 5.19%. This is the expected effect on the exempt group due to bunching alone, provided the responses of the bunchers in the exempt and non-exempt groups are the same. The second column contains the difference-in-difference estimates for the non-exempt group, which are the observed effects. By subtracting the implied bunching effect in the first column from the observed effect in the second column, I obtain the implied effect due to the extensive-margin response, in the final column.