Journal of Public Economics 152 (2017) 133–142

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Optimal tax administration☆

Michael Keen [a], Joel Slemrod [b,*]

[a] Fiscal Affairs Department, International Monetary Fund, United States
[b] Department of Economics, University of Michigan, United States

ARTICLE INFO

Article history:
Received 26 September 2016
Received in revised form 26 April 2017
Accepted 26 April 2017
Available online 9 May 2017

JEL classification:
H21
H26

Keywords:
Tax administration
Tax compliance
Optimal taxation

ABSTRACT

This paper sets out a framework for analyzing optimal interventions by a tax administration, one that parallels and can be closely integrated with established frameworks for thinking about optimal tax policy. Its key contribution is the development of a summary measure of the impact of administrative interventions—the "enforcement elasticity of tax revenue"—that is a sufficient statistic for the behavioral response to such interventions, much as the elasticity of taxable income serves as a sufficient statistic for the response to tax rates. Among the applications are characterizations of the optimal balance between policy and administrative measures, and of the optimal compliance gap.

1. Introduction

This paper sets out some simple analytics of optimal tax administration, focusing on three questions.

First, and perhaps most fundamentally: What exactly do policy makers need to know about the costs and effectiveness of administrative interventions in order to set them optimally? The analogous question on the policy side has received considerable attention, with much of the recent literature focusing on the circumstances in which the elasticity of a reported tax base with respect to the net-of-tax statutory rate[1] is a sufficient statistic for the optimal choice of marginal tax rate. An extensive empirical literature focuses on estimating this elasticity, particularly for income taxes, in which setting it is known as the elasticity of taxable income.[2] But what of the administration side? The characterization of optimal interventions has certainly received some attention (starting with the classic treatment in Mayshar (1991)), but leaves open the question of whether there are concepts analogous to the taxable income elasticity that might prove equally useful in guiding empirical work on the proper extent and design of administrative interventions.

The second question is at the heart of much practical policy debate. Policy makers facing a need to raise more revenue have broadly two alternatives: to raise rates, or to take potentially costly measures to improve compliance. Rhetoric on this abounds, but theory has provided little guidance on this most basic of policy choices: Is it better to raise an additional dollar of revenue by increasing statutory tax rates or by strengthening tax administration so as to improve compliance? Closely related to this is the question of how constraints on one dimension of tax system design affect the optimal choice of the other: if, for instance, tax administration is weaker than would ideally be the case, does that call for higher tax rates than would otherwise be optimal, or for lower?

The third question concerns the significance and use of the concept of the 'compliance gap:' the difference between the amount of tax legally due and that actually collected. This is an intuitively appealing indicator of the effectiveness of a revenue administration—with clear advantages, for instance, over the simple comparison of cost-revenue ratios (i.e., administration (and/or compliance) costs relative to revenue raised) that has traditionally been a focus in assessing the performance of tax administrations. Reflecting this appeal, the calculation and analysis of compliance gaps has become a major focus of effort in the last few years—and interest continues to grow: they are now regularly produced for a range of taxes in the U.S. by the Internal Revenue Service (IRS, 2012) and in the U.K. by HMRC, for example, and for the VAT in the member states of the European Union.[3] While many technical questions

☆ We thank Anne Brockmeyer, John Creedy, Vitor Gaspar, Norman Gemmell, Christian Gillitzer, Shafik Hebous, Christos Kotsogiannis, Ben Meiselman, Alan Plumley, Mick Thackray, the referees and the editor for many helpful comments and suggestions, and Will Boning for excellent research assistance. Views and opinions are ours alone, and should not be attributed to the management or Executive Board of the International Monetary Fund.

* Corresponding author.
E-mail addresses: mkeen@imf.org (M. Keen), jslemrod@umich.edu (J. Slemrod).

[1] That is, unity minus the tax rate.

[2] An extensive review of the literature on the elasticity of taxable income is provided by Saez et al. (2012).

[3] See HMRC (2015) and European Commission (2015) and, on tax gap analysis more generally, Keen (2013) and IMF (2015).

http://dx.doi.org/10.1016/j.jpubeco.2017.04.006
0047-2727/© 2017 Published by Elsevier B.V.

arise in measuring and analyzing compliance gaps, a more fundamental criticism is leveled by Gemmell and Hasseldine (2014). They stress that their mechanical construction abstracts from behavioral responses, which means, for instance, that measures which reduce the compliance gap may also reduce real activity and hence perhaps even revenue: an activity may be privately profitable if the associated tax collected is below that legally due, but not if that tax is fully remitted. The welfare impact of administrative interventions thus cannot be inferred simply from associated changes in the compliance gap. Given too the costs of implementing such interventions, for both government and taxpayers, it is clear that—while much analysis often proceeds on the implicit presumption that, whatever the compliance gap is, it is too big—the optimal compliance gap is not zero. The question then arises: What exactly is the 'optimal' compliance gap? More generally, how can we know whether an observed compliance gap is too big or too small?

At the heart of the answers to these questions set out in this paper is the concept of the *enforcement elasticity of tax revenue*: the responsiveness of revenue collected to administrative interventions (one such elasticity, in principle, for each instrument of administration). This, the analysis shows, is the administration-side analogue of the elasticity of taxable income, acting, in the same way, as a sufficient statistic for the behavioral impact of administrative interventions that encompasses effects on the levels of both true and concealed activities. At an optimum, the enforcement elasticity is equated to a straightforward variant of the usual cost-revenue ratio, a simple rule that also provides a clear role for a quantity that, with little theoretical rationale, has long been a center of attention in the traditional literature on (and practice of) tax administration (dating back at least to Sandford (1973)). The choice between policy and administrative measures, the analysis further shows, turns on the balance between these two elasticities, along with an even simpler form of the cost-revenue ratio. And, on the third question, it is shown that the optimal compliance gap is characterized, in a benchmark case, by a simple inverse elasticity rule, the relevant elasticity in this case being that of evasion with respect to enforcement; more generally, the optimal gap also reflects the distinct reactions to administrative actions of both real earnings (as stressed by Gemmell and Hasseldine (2014)) and (legal) avoidance.

Section 2 establishes core results in a simple benchmark case,[4] which section 3 then generalizes. Section 4 sets out some extensions, and Section 5 discusses the empirical application of the theoretical structure developed. Section 6 concludes.

2. Analyzing tax administration: a simple framework

To start with a simple and standard case (along lines similar to Chetty (2009) and Slemrod (2001))—taking this as a metaphor for wider circumstances in which both earning and concealing from the authorities some tax base are costly— consider a representative individual with quasi-linear preferences (this being the first of several assumptions relaxed in the next section) of the form

$$W = x - \psi(l) + v(g), \tag{1}$$

where x denotes private consumption, l hours worked and g public spending from which the consumer directly benefits; ψ and v are both strictly increasing and, respectively, strictly convex and concave. Public spending is financed by a proportional tax on income at rate t,[5] so that consumption is given by

$$x = wl - t\cdot(wl - e) - c(e,\alpha), \tag{2}$$

where w is the (exogenous) wage rate, e the amount of income not revealed to the tax authorities (so that $z \equiv wl - e$ is taxable income[6]), c denotes the private costs associated with that failure to reveal (discussed further in a moment), and α—central in what follows—is some continuously variable enforcement parameter[7] that is at the control of the tax administration. This last is defined so that $\alpha > 0$, with higher values of α increasing both the private costs of inaccuracy in revealing income and the marginal cost of inaccuracy, so that $c_\alpha > 0$ and $c_{e\alpha} > 0$.[8]

For the present, we take the enforcement instrument α to be a single continuous variable, such as the probability of audit or the ease of remitting payment; the case of multiple instruments is considered in Section 4. Many tax reforms, however, are not readily characterized as a continuous parameter change. This is true not only of discrete reforms among a single parameter dimension (a large increase in audit activity, for instance) but, more fundamentally, qualitative ones: the adoption of a large taxpayer unit, for instance, or the movement from a tax-type to a functional structure of the tax administration. The approach set out here is readily extended to deal with large and/or qualitative reforms by using the discrete analogue to the differential analysis below. We do not elaborate on the details of doing so here,[9] but illustrate the point with a practical application in Section 5.

We refer to c interchangeably as costs of compliance or of concealment,[10] and for brevity simply assume throughout that $c_e > 0$, $c_{ee} > 0$, and that the taxpayer chooses strictly positive concealment.

Substituting Eq. (2) into Eq. (1), with g exogenous the individual's choices of hours worked and concealment are characterized, respectively, by the necessary conditions $(1-t)w - \psi'(l) = 0$ and $t - c_e(e,\alpha) = 0$ (so that, as just noted, $c_e > 0$ in equilibrium). These define solutions $l(t, w)$ and $e(t,\alpha)$, with $e_\alpha = -c_{e\alpha}/c_{ee} < 0$ and $e_t = 1/c_{ee} > 0$. Note that labor supply is here independent of enforcement, reflecting the independence of concealment costs from true income—a feature that will be relaxed in the next section.

The administrative costs associated with the intervention α are denoted by $a(\alpha)$, with $a' > 0$. In many contexts, it will be natural to measure an intervention by its cost, so that $a(\alpha) = \alpha$; indeed doing so simplifies many of the expressions that follow, and we shall explore this formulation below.[11] The general formulation $a(\alpha)$, however, has several advantages. It brings out most clearly an instructive parallel with the familiar elasticity of taxable income, and shows more sharply the symmetries and differences between administration and

[4] Creedy (2016) provides a graphical exposition of this and other propositions presented here, and illustrates them using explicit functional forms for the key relationships. Zoutman and Bas (2016) pursue a related objective, extending the Mirrlees (1971) model of optimal non-linear income taxation with a monitoring technology that allows the government to verify labor supply at a positive, but finite, cost. They characterize the joint determination of monitoring and the tax rate schedule, and show that the optimal intensity of monitoring increases with the marginal tax rate and the labor supply elasticity. Best et al. (2015) develop a model that characterizes the optimal tax system with sufficient statistics that include an evasion elasticity, although their focus is the tradeoff between production efficiency and revenue efficiency.

[5] There is, thus, no demogrant; the implications of nonlinear taxation are considered after Proposition 1 below.

[6] We follow the standard usage here: strictly, taxable income is wl.

[7] Describing administrative interventions in terms of enforcement might seem to neglect the encouragement of voluntary compliance that many tax administrations see as a core part of their work. But this usage is for brevity only, and the framework here can be interpreted to encompass, for example, not only what Alm (2014a) calls the 'enforcement paradigm' of tax administration but also his 'service' and 'trust' paradigms. Measures which encourage voluntary compliance, for example, such as prepopulating returns, can be seen as increases in α that make it easier to be honest (equivalently, harder to be dishonest), so that $c_{e\alpha} > 0$, just as is assumed below.

[8] Derivatives are indicated by subscripts for functions of several variables, and by primes for functions of just one.

[9] This is done in the working paper version of this paper, Keen and Slemrod (2017).

[10] The former is the more familiar term in the literature by which we will later calibrate some of our results, and calls to mind the costs that may be incurred by the effort to reveal taxable income with full accuracy, so that $c_e < 0$. The latter term calls to mind the costs incurred in hiding income from the tax authorities in which case $c_e > 0$. Both concerns can be accommodated within the general formulation here. It might be, for instance, that c is U-shaped in e, with some least-cost level of under-declaration relative to which it is costly to be either less or more dishonest. No taxpayer, however, will under-declare at a point at which $c_e < 0$, because a little more concealment would then reduce both tax paid and non-tax costs c.

[11] See Eq. (28) and the discussion thereabouts.