WP 2014-11
March 2014

Working Paper

Charles H. Dyson School of Applied Economics and Management
Cornell University, Ithaca, New York 14853-7801 USA

THRESHOLDS, INFORMALITY, AND PARTITIONS OF COMPLIANCE

Ravi Kanbur and Michael Keen

It is the Policy of Cornell University actively to support equality of educational and employment opportunity. No person shall be denied admission to any educational program or activity or be denied employment on the basis of any legally prohibited discrimination involving, but not limited to, such factors as race, color, creed, religion, national or ethnic origin, sex, age or handicap. The University is committed to the maintenance of affirmative action programs which will assure the continuation of such equality of opportunity.

Thresholds, Informality, and Partitions of Compliance

Ravi Kanbur[*] and Michael Keen[**]

October 14, 2013
Revised February 2014

Abstract: This paper explores the impact and design of tax thresholds, and the light this casts on notions of 'informality.' It shows how thresholds generate partitions of the population of potential taxpayers by different forms of compliance and non-compliance. The richness of these partitions means one should resist thinking of 'the informal sector' as an undifferentiated mass, but instead recognize that there are quite distinct varieties of informality with potentially very different policy implications. We characterize the forces shaping such taxpayer partitions, and, within that setting, the optimal threshold and the partitions it induces. The analysis is extended to the (realistic) case in which taxpayers face multiple tax (or other) obligations, showing how a threshold on one obligation affects partitioning and optimal threshold choice with respect to the other.

Keywords: Informality, tax compliance, tax threshold

JEL: H21, H26, Q17, Q23

Acknowledgements: We are grateful to three referees, Philippe Wingender, and seminar participants at the IIPF Congress and the IMF for many helpful comments and suggestions, and to Kelsey Moser for excellent research assistance. The views expressed here are ours alone, and should not be attributed to the IMF, its Executive Board, or its management.

[*] Cornell University, 301J Warren Hall, Ithaca, NY 14853-7801; sk145@cornell.edu
[**] Fiscal Affairs Department, International Monetary Fund, Washington DC, 20431; mkeen@imf.org

I. Introduction

This paper aims to establish and explore the links between two threads in the public finance literature. One is the use of tax thresholds to partition taxpayers[1] into those who are liable to pay tax and those who are not. The other is the notion of 'informality' as a central challenge for tax design and implementation, especially, but not only, in developing countries. The two are intimately related because making sense of 'informality' as a concept requires recognizing that it subsumes a range of non-compliant (and compliant) behaviors with the tax threshold being a primary determinant of those behaviors; and so, by the same token, the setting of thresholds needs to recognize the potentially complex impact on the extent and nature of 'informality.' Using a highly stylized model to formalize these ideas, the central analytical purpose of the paper is to characterize partitions of taxpayers by different types of compliance and non-compliance, show how this is shaped by the level of the threshold, and explore the implications for the optimal choice of threshold[2] and associated partition—and hence, more loosely, for optimal informality.

Most taxes (though not all) have some threshold below which no tax is payable, the fundamental rationale being simply that costs of administration (to the tax authorities) and of compliance (to the taxpayer) argue against imposing taxes on the smallest taxpayers. These thresholds can be of two types. They may involve a 'kink', in the sense that while the marginal tax rate switches from zero to strictly positive as the threshold is crossed, tax is imposed only on the amount by with the base (which we shall call 'income') exceeds that threshold; tax liability itself is thus continuous around the threshold.[3] Almost all personal income taxes are of this familiar form. The other type of threshold is a 'notch',[4] the difference being that in this case tax is charged once the threshold is exceeded not just on the excess over the threshold but the full amount of the base: tax liability thus jumps discontinuously at the notch. Most value added taxes (VATs), for instance, have this feature: once a trader's sales exceed some threshold level of turnover, they are required to charge VAT on all of their sales. And there are many other if less obvious examples of such jumps in liability: simplified income tax regimes may apply, for instance, only below some income (or other) threshold. Indeed Slemrod (2010) shows that such 'notch' structures are a much more pervasive feature of tax (and other) systems than is often recognized. Importantly too, even if the formal threshold itself is in the form of a kink, the compliance costs associated with paying some rather

[1] We use 'taxpayers' as shorthand for the universe of firms or individuals (not restricting it to those who actually do or should pay tax).

[2] There are other considerations in the optimal choice of threshold not addressed here. It may for instance be desirable to tax smaller firms, even if the costs of doing so exceed the revenue raised, in order to reduce the distortion of competition between taxed and untaxed firms (Keen, 2013). Some also see political economy benefit in levying some charge on even the smallest taxpayers as a way of encouraging them to hold policy makers accountable (see for example OECD (2008)). We abstract from these considerations here.

[3] It is kinks that are exploited in the seminal work of Saez (2010) on the estimation of elasticities of taxable income.

[4] Arrangements of this kind are sometimes also referred to as 'slab' or 'cliff' structures.

than no tax are likely best thought of as including some fixed component that, for the taxpayer, has the same effect as a tax notch [5]

The focus in this paper is on the notch form of threshold. This reflects not only their practical importance, but also the sharpness with which they raise the key compliance issues with which we are concerned and, not least, the ease of comparison with previous work. The impact and design of notch thresholds has received some previous attention. Kleven and Waseem (2013) use such a feature found (very unusually) in the personal income tax of Pakistan to identify the elasticity of taxable income, for instance, while Chatterjee and Wingender (2011) look at a similar feature of indirect taxes in India. Closer to current concerns, however, is work on the optimal choice of threshold of Dharmapala, Slemrod and Wilson (2011), Zee (2005) and—used as a benchmark in the analysis below—Keen and Mintz (2004). All this literature has assumed, however, that all taxpayers are fully compliant. But non-compliance with tax laws is a universal concern: taxpayers might, in particular, falsely declare under the threshold (or simply not report), or declare above the threshold but below their true liability.

One key task here is to extend previous analyses of tax thresholds to allow for such non-compliance. What partitions, in terms of different behavioral responses, does a notch threshold induce? How do these partitions of tax compliance change when the threshold changes? And what is the optimal threshold? A useful reference point for this last question is provided by a simple formula for the optimal VAT threshold derived by Mintz and Keen (2004) for the case in which not only are all compliant, but their outputs are fixed and perfectly observable to the authorities, the natural question, addressed here, then being: Recognizing these different types of compliance and non-compliance, is the optimal threshold higher or lower than that simple rule implies?

One other and previously neglected aspect of threshold choice is also taken up here. Such attention as the issue has received presumes that firms face only one obligation and, so, only one threshold. In practice, they often face multiple obligations and, correspondingly, may well face multiple thresholds. This may be true simply in tax terms—distinct thresholds for the VAT and a presumptive income tax, for instance—but can also arise because firms also face non-tax obligations: the requirements of labor law, for instance. The question then is how a threshold in relation to one obligation affects partitions in respect of another, and how that in turn affects the optimal levels of the various thresholds.

These questions of threshold design and impact also speak to the concept of 'informality' as it is used in the public finance literature—the second thread with which we are

[5] It is an essentially universal empirical finding that compliance costs relative to some indicator of size tend to fall with that indicator, and quite sharply at smaller sizes, so that any reasonable approximation of compliance costs will include a substantial fixed component. See, of many possible examples, European Commission (2004) and, for developing countries, Coolidge (2012). There is of course a large and growing literature on different types of tax or regulatory regimes which induce kinks and discontinuities of different types in incentive structures: see for example, in addition to Saez (2010), Bastani and Selin (2012), Almunia and Lopez-Rodriguez (2013) and Garicano, Lalarge and van Reenen (2013).

concerned.[6] Accounts of tax challenges in developing economies commonly put informality at or near the top of the list. Auriol and Warlters (2005), for example, summarize the difficulties enumerated by Burgess and Stern (1993) as the single point that "developing countries have large informal sectors that are difficult to tax"; Keen and Simone (2004) similarly highlight that "informality is extensive," and "[t]axing the informal economy" leads the African Development Bank's list of policy priorities (Mubiru, 2010). The term is prominent in policy discourse in developed countries too: IMF (2013) refers to the "large size of the informal economy" in Greece as a source of "low revenue efficiency." But while it is easy find statements of the importance of 'informality' for tax design and implementation, it is very hard to find clear definitions of what exactly it is.

The technical literature tends to model 'informality' simply as non-remittance of tax. But there are many reasons why tax might not be remitted, a point that—linking the two themes of the paper—thinking about tax thresholds makes abundantly clear. A threshold creates an obvious distinction between those who end up above the threshold and those who end up below. But there also marked and, as will be seen, potentially policy-relevant and threshold-driven differences within each of these groups. Those who end up remitting no tax, for instance, may be of three types: those whose maximum potential income is below the tax threshold; those whose potential income exceeds the threshold but who, perfectly legally, adjust down to just below the threshold in order to avoid tax; and those whose actual income is above the threshold but who falsely and illegally claim to have income below the threshold. These three types of behavior evidently have very different implications for output, welfare and policy design: it is not simply the fact of non-remittance of tax that matters, but the reason it occurs. Notions of informality that miss such distinctions risk muddled analysis.

Elaborating on this, a central argument of this paper is that lumping various categories of compliant and non-compliant tax behavior under the single label 'informality,' trying to summarize it in a single quantitative measure, and presuming that 'it' is a problem is both analytically problematic and potentially highly misleading for policy purposes. What is important is to recognize the varieties of informality and how each responds to policy instruments. It will be seen, for instance, that the setting of optimal thresholds may well involve balancing a reduction in some types of informality against an increase in others—and perhaps even increasing, on some measures, its overall extent. Recognizing the multiplicity of obligations that taxpayers commonly face, with their associated

[6] The perspective that we take on informality in this paper is thus very much tax-driven. There are of course others. There is, in particular, a long-standing interest in informality from a labor perspective: see the discussion in Kanbur (2009) and Chatterjee and Kanbur (2013). These differing perspectives can lead to quite different notions of 'informality.' For instance, OECD (2009), focusing on employment aspects, speaks of "Shoe shine workers in Cairo, street vendors in Calcutta: this is informal employment." That is perfectly reasonable from a labor market perspective; from a public finance perspective, however, it is far from clear that one would actually want these groups to charge VAT or remit income tax, and they may indeed be fully (or almost) tax compliant. Others will think of informality spanning a range of aspects of behavior, associated for instance with limited book-keeping, heavy reliance on cash transactions and the like. All this creates further difficulties and imprecision in the use of the term.

thresholds, will be seen to point even more strongly to the need for a more subtle and holistic approach to notions of informality, compliance and partitioning.

The plan of the paper is as follows: Section II sets out a simple model that enables an endogenous partitioning of taxpayers in terms of compliant and non-compliant behaviors, and explores how this partitioning depends on the level of the threshold. It also applies these results to the issues in conceptualizing informality just raised. Section III then characterizes the optimal threshold, again drawing implications for understanding and addressing the varieties of informality. Section IV extends the analysis to the case of two tax regimes, each with its own possible threshold, analyzes cross-effects on compliance across the two, and considers whether or not it is optimal to have a common threshold for both. Section V concludes.

II. A Partition of taxpayers

This section sets out a framework enabling a unified treatment of different forms of compliance and non-compliance in the presence of a notch threshold for tax liability.

A simple model

We consider a population of individuals differing in the maximum potential income[7] they can earn (without effort or any other costs), denoted *Y*, which is taken to be exogenous and distributed continuously with density $f(Y)$ and twice differentiable distribution function $F(Y)$ on support $[0, \infty)$,and population size normalized at unity.[8] They can, if they wish, choose to earn any amount less than this; they can also choose what income to reveal to the government. Their objective in all this is simply to maximize their income after tax and other costs.

The tax system is of the 'notch' form: income above some threshold Z is taxable in its full amount (not just that part above *Z*). The role and, later, choice of this threshold will be the central focus in what follows. Those remitting any strictly positive amount of tax, it is assumed, also incur fixed compliance costs of $K > 0$. In practice, of course, compliance costs are likely to increase with the extent and complexity of a taxpayer's business. Taking them to be the same for all is a simple way of capturing the increasingly well-documented stylized fact noted at the outset, that these costs decrease markedly relative to the scale of activity. The marginal tax rate T is assumed constant,[9] so that the net income of a fully compliant taxpayer with income above the threshold is $(1 - T)Y - K$.

[7] For definiteness, the discussion is for the most part cast in terms of 'individuals' or 'firms' and 'income,' but clearly many interpretations are possible.

[8] This is in the spirit of related treatments in which entrepreneurs differ in ability and make some costly input choice (such as Dabla Norris, Gradstein and Inchauste (2008) and Keen and Mintz (2004)). This approach would lead to cut-offs of the kind discussed below but defined in terms of ability rather than potential income. The simpler representation of output decisions here facilitates analysis of the somewhat involved choice between the multiple regimes about to be described.

[9] This of course is in order to abstract from distinct effects arising from the shape of the tax schedule. For developing countries, there is evidence—now rather dated, but still plausible (Gautier and Gersovitz

While for clarity we speak of Y here as 'income,' most income taxes, of course, are not of this notch form, but rather levy tax only on the amount $Y - Z$ above the threshold, not on Y itself. Almost all value added taxes, however, do have precisely this notch feature, and the structure here is readily re-interpreted in that way.[10] The rationale for considering a notch structure is not realism in describing income taxes but rather to capture the pervasiveness of notches in tax codes (and in regulatory ones too), stressed by Slemrod (2009), and to enable later results to be related to earlier analyses of taxpayer partitioning under much simpler circumstances in Keen and Mintz (2004). It is important too to note that while the hallmark of the notch tax structure is that it introduces a discontinuity in net income when the threshold is crossed, fixed compliance costs induce such a discontinuity even in the absence of a notch in the tax structure itself.

Faced with this tax system, individuals choose between four types of behavior: they may earn and declare their true potential income (and be above or below the threshold accordingly); they may genuinely reduce their income to below the threshold; they may dishonestly opt out of the tax system entirely; or they may earn their full potential income but declare less. To focus on the extensive margin of the choice between these behaviors, we effectively abstract from the intensive margin within each: that is, once the choice of regime is made, there is no substantive decision to be made as to true or declared income.

Fully honest taxpayers in this setting are, thus, potentially, of three types. There are those with maximum potential income below the threshold, who declare truthfully and pay no tax—this is what one might expect of 'micro' traders, so we refer to these as type *M*. There are those who truthfully declare above the threshold—type *L*. And there are those whose maximum income is above the threshold but who choose to save tax and compliance costs by reducing their income and truthfully declaring just below the threshold[11]—we call these *adjusters*, type *A*. The existence of this last option means that there will be a 'hole' above the threshold: a range of income over which—as a consequence of both fixed compliance costs and the notch tax structure—adjusting to below the threshold leads to higher net income.

Types *M*, *L*, and *A* are all, in different ways, fully compliant. But we also allow for the two different forms of non-compliance mentioned above.

One is to opt out of the tax system by either becoming invisible to the tax authorities—a 'ghost'—or by falsely declaring just under the threshold. This though, we assume, involves some real cost,[12] perhaps through the need to earn or hide income in convoluted

---

(1997), Gauthier and Reinnika (2001))—that larger firms benefit more from exemptions, which would reinforce the pattern of compliance derived below.

[10] In this case (and ignoring input costs) Y would be interpreted as potential sales revenue, with the assumption that this is independent of the tax system corresponding to an assumption that the consumer price is fixed—as would be the case, for instance, in a small economy open to trade in the commodity/ies of interest.

[11] To keep things manageable, we preclude the possibility of splitting activity between more than one unit below the threshold (the legality of which will depend on the grouping rules for the tax in question).

[12] For the welfare part of the analysis, it is assumed that these costs imply a real reduction in income, and are not simply a transfer.

ways so as to escape notice, so that net income is in this case reduced to $(1-\gamma)Y$—we call these type *B,* for *bounders*—capturing both their ungentlemanly behavior and the notion that they may be falsely presenting their income as below the lower bound for liability.[13] It is assumed throughout that $1 > \gamma > T$: without this, there could be no honest taxpayers.

The other type of non-compliance is to declare only some proportion $\lambda \in (0,1)$ of true income—the same proportion, we assume, for all taxpayers—incurring, in doing so, some total cost $\Gamma(Y)Y$, with $\Gamma(Y)$ non-negative, increasing, convex and differentiable.[14] So net income in this case is $(1-\lambda T-\Gamma(Y))Y-K$. We refer to such taxpayers as *cads*—another label of dishonor in Victorian England; type *C*. The costs they incur may reflect, for instance, expected penalties, so that Γ could also depend on λ and T (though this need not be made explicit since both will be held constant in what follows); they might also capture bribes paid to tax inspectors, or the costs involved in (legal) avoidance. The assumption that the possible degree of under-reporting λ is the same for all is restrictive, of course, and has some implications that should be noted. One is that this it may imply that some taxpayers declare an income in the 'hole' above the threshold mentioned above, so that, were that declaration honest, they would be manifestly better off adjusting to below the threshold—which would make them prime targets for an active audit strategy. As a practical matter, this may not be too unrealistic: when income is subject to some uncertainty, honest taxpayers can indeed find themselves in this range and tax authorities do not seem to look particularly intensely at those declaring just above the VAT threshold. More awkward is that mechanical application of the rule can result in declared incomes that are below the threshold but nonetheless taxed; this, however, can be ruled out by parametric restrictions that do not involve *Z*.[15]

Partitioning taxpayers

Individuals choose whichever of these options gives them the highest net income. This choice can be characterized in terms of various cut-off levels of income.

Bounding will be strictly preferred to adjusting, for instance, if and only if

$$(1-\gamma)Y > Z \tag{2.1}$$

which defines a cut-off level of income

[13] Note the assumption here that the net income of those who entirely vanish from the tax system is independent of the threshold (one implication being that this category of behavior cannot plausibly be interpreted as including artificial splitting into more than one unit below the threshold)

[14] Not all models of concealment imply an increasing $\Gamma(Y)$: in the model of corrupt tax inspection in Hindriks, Keen and Muthoo (1999), for example, the relationship between the equilibrium bribe and true income has an inverse-U shape. But nor is the present assumption entirely implausible: an increasing Γ is implied, for instance, if concealment costs are quadratic in the amount concealed. The general approach to modeling concealment costs here is in the spirit of, and discussed further in, Slemrod (2001).

[15] For instance: adjusting is preferred to concealing at all income levels below θ_{AC} defined in (3.11) below. Clearly $\theta_{AC} > Z/(1-\lambda T)$, so that a sufficient condition for θ_{AC} to exceed the highest income level, Z/λ, at which mechanical under-declaration would put the taxpayer under the threshold is that $\lambda > 1/(1+T)$.

$$\theta_{AB}(Z) = \frac{Z}{1-\gamma} \tag{2.2}$$

such that all those with income above θ_{AB} prefer bounding to adjusting, and all those with lower income prefer the opposite. (The general notation introduced here is that θ_{JK} denotes the level of income at which net income under behaviors *J* and *K* give the same net income). Similarly, those (having income above *Z*) with income θ_{BC} defined by

$$\big(1 - \lambda T - \Gamma(\theta_{BC})\big)\theta_{BC} - K = (1-\gamma)\theta_{BC} \tag{2.3}$$

are indifferent between concealing and bounding, while those with income θ_{CL} such that

$$(1-T)\theta_{CL} - K = \big(1 - \lambda T - \Gamma(\theta_{CL})\big)\theta_{CL} - K, \tag{2.4}$$

or, equivalently, such that

$$(1-\lambda)T = \Gamma(\theta_{CL}), \tag{2.5}$$

are indifferent between honesty and concealing. From (2.3) and (2.4), those with incomes above θ_{BC} will prefer concealing to bounding, and those above θ_{CL} will prefer honesty to concealing, so long as, over the relevant range,

$$\frac{\partial(\Gamma(Y)Y)}{\partial Y} \in ((1-\lambda)T, \gamma - \lambda T), \quad \forall Y, \tag{2.6}$$

which we henceforth assume to be the case.[16] It will also be assumed throughout that:

$$(\gamma - T)\theta_{BC} < K\,, \tag{2.7}$$

which means that those indifferent between bounding and concealing prefer bounding to full compliance. The restrictions in (2.6) and (2.7) are of course somewhat arbitrary. But it would be tedious to consider all conceivable permutations, and these assumptions serve to focus attention on what seem to us likely to be the most relevant possibilities for practical policy design.

While this covers only three of the binary comparisons that individuals need to make, it is enough to begin analyzing how they will be partitioned in equilibrium. This will clearly depend on the threshold *Z*, and indeed a key part of our task here is to explore precisely that dependence. To begin, suppose first—a temporary assumption, relaxed later—that the threshold is low enough and/or compliance costs high enough, that

$$Z < \big(1 - \lambda T - \Gamma(\theta_{BC})\big)\theta_{BC} - K \tag{2.8}$$

[16] The interval on the right of (2.6) is larger, and the assumption in that sense more plausible, the greater is the excess of the proportionate loss of income from bounding, γ, over the tax rate *T*.

(so that those indifferent between bounding and concealing prefer the latter to adjusting). This provides a useful benchmark case in which all three types of behavior arise in equilibrium:

PROPOSITION 1: *Given (2.8), taxpayers partition themselves as:*[17]

(i) Those with $Y < Z$ *or* $Y \geq \theta_{CL}$ *declare honestly: these are types* M *and* L *respectively.*

(ii) Those with $Y \in (Z, \theta_{AB})$ *adjust out of the tax system: type* A.

(iii) Those with $Y \in [\theta_{AB}, \theta_{BC})$ *are bounders: type* B.

(iv) *Those with* $Y \in [\theta_{BC}, \theta_{CL})$ *are cad: type* C.

Proof: It suffices to show that $\theta_{CL} > \theta_{BC} > \theta_{AB}$. For the first of these inequalities, substituting the implication of (2.5) that $\lambda T = T - \Gamma(\theta_{CL})$ into (2.3) and rearranging gives

$$[\Gamma(\theta_{CL}) - \Gamma(\theta_{BC})]\theta_{BC} = (T - \gamma)\theta_{BC} - K\,. \tag{2.7}$$

Since Γ is increasing, the result then follows from (2.7). For the second inequality, suppose to the contrary that $\theta_{BC} < \theta_{AB}$. Then, from (2.2), $(1-\gamma)\theta_{BC} < Z$ and (2.3) implies

$$\big(1 - \lambda T - \Gamma(\theta_{BC})\big)\theta_{BC} - K < Z, \tag{2.8}$$

which violates (2.8).

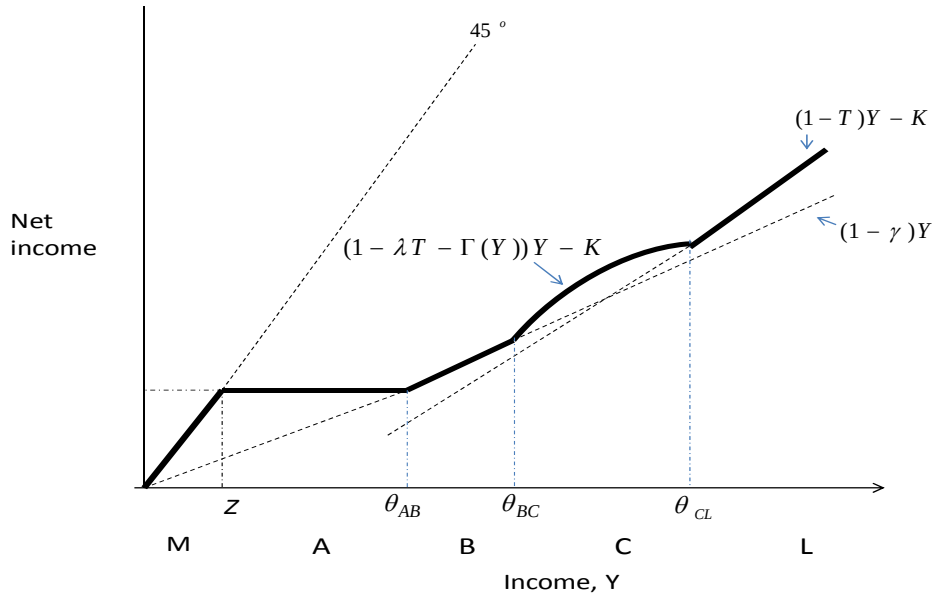
The outcome, given a threshold as in (2.8), is thus as in Figure 1, with taxpayers' choices leading to a sharp partitioning by size in which all forms of compliance and non-compliance arise over some range. Those with the highest potential incomes are fully compliant and do not adjust their real behavior; those just below conceal some of their income, those below that opt out of the tax net, the next adjust out, and the very smallest do not react at all to the tax system.

The neatness of this partitioning of course reflects the underlying parameter assumptions, and one would not expect compliance behaviors to be so mechanically associated with potential income in practice, but to be shaped also by various forms of heterogeneity. The pattern is, nonetheless, broadly plausible and will likely have a ring of truth with practitioners. While we know of no evidence allowing the identification of all such types of behavior in relation to a single tax, there is increasing evidence of their potential importance. Chatterjee and Wingender (2012) and Kleven and Waseem (2013), for instance find bunching below notches in India and Pakistan respectively, while Onji (2009) detects VAT threshold effects on the distribution of firm size in Japan. The extent of 'bounding' is less well-documented. Some of this will be reflected in the bunching just described, as false declarations below the threshold, but in other cases it will simply

[17] As a tie breaker, we assume that those indifferent between regimes select that preferred by those with slightly higher incomes.

result in taxpayers disappearing from sight. Thus Erard and Ho (2001) find that about 7 percent of the liable population in the U.S. fail to file, while results for Cameroon in Gauthier and Gersovitz (1997) imply that about 15 percent of

Figure 1: A partition of taxpayers by compliance



the firms in their sample were tax evaders who had had no contact with the tax authorities. There is thus good reason to suppose that all five forms of response do indeed occur, and can be significant.

Implications

One consequence of the compliance partitioning of taxpayers into the five groups shown in Figure 1 is a nuancing of the notion of a 'missing middle' in the distribution of firm size, which has long seen as a feature of developing counties in particular (see for instance Tybout, 2000). The variety of behaviors displayed here means that this idea needs to be interpreted with some care, and a distinction made, in principle, between three distinct distributions of income: the true, that reported to the tax authorities, and that which may be reported by other agencies (through production surveys, for instance).

There is clearly, for instance, a 'hole' above the threshold—in all three distributions—in the sense that all those with potential incomes mass between Z and θ_{AB} choose (perfectly legally) to reduce their incomes to (just below) the threshold. But that is not the end of the story. Some bounders may choose to earn (but not declare to the tax authorities) incomes lying within this range, going some way towards filling up the hole in the distribution of true incomes; but others are likely to be either entirely absent from

recorded data or to falsely amplify the mass observed below the threshold. At the same time, some of those with potential incomes above θ_{BC} may (falsely) report incomes within the region of potential incomes associated with adjusting or bounding. Indeed Kleven and Waseem (2103) find that half or more of their sample of self-employed filers in Pakistan declare in dominated ranges over which (even ignoring the saving in compliance costs) adjusting out of the system would increase net income (though their interpretation is in terms of optimization frictions rather than concealment). The varieties of non-compliance thus imply a series of subtleties for tax-based approaches to thinking about and identifying a 'missing middle,' as well as understanding the various bunches and holes, true and false, that might arise.

More directly related to the concerns raised in the Introduction is the perspective that the partitioning in Figure 1 provides on ideas of informality. Who, to begin with, might one call 'informal' in this partitioning? Those in group *L* are both liable to tax and fully compliant, so they would presumably be judged to be 'formal.' Taking 'informality' to be the complement of formality, those in groups *M*, *A*, *B* and *C* would then be the 'informal.' Thus defined, informality embraces quite distinct types of both non-compliance (bounders and cads; B and C) and compliance (micro and adjusters; M and A).

What is of particular interest here is how the extent and nature of this informality is affected by the tax threshold *Z*. Strikingly, since (recalling (2.5)) the critical boundary θ_{CL} is independent of *Z*, a (small) change in the threshold has no effect on the overall *number* of informal individuals—in the sense that it has no impact on the total number in groups *M*, *A*, *B* and *C*. What it does affect, however, is the *composition* of that aggregate informality: as is evident from Figure 1, a small increase in *Z* will increase the number of adjusters but reduce the number of bounders by exactly the same amount. And that, as will be seen, generally means a change in the overall *output* of informal firms. Similarly, since (from (2.3)) a small change in the threshold has no effect on θ_{BC}, it has no effect on tax revenue. But such a change does affect welfare, nonetheless, because of the induced change in private output. The composition of informality—the balance, that is, between different forms of compliance and non-compliance—matters for policy design.

All of this raises the question, to which we now turn: Given the complexity of its effects, what is the optimal level of the threshold *Z*, and what does that imply for the optimal structure and nature of 'informality'?

III. Optimal Thresholds

To explore the optimal choice of threshold—and the optimal partitioning of taxpayers it implies—we now dispense with the assumption in (2.8). Expositionally, however, it proves convenient to begin the analysis by imagining that the threshold is initially set at a level which satisfies (2.8).

Leaving equity considerations aside, the object of policy is taken to be the maximization of the sum of private incomes and net tax revenue, with the latter weighted by some

factor $\delta > 1$ that can be thought of as the marginal cost of public funds raised from other instruments available to the government, or simply as the marginal value of public spending. Introducing now fixed costs of administration, incurred by the government for all those remitting tax,[18] denoted by *N*, social welfare, given the partition in Proposition 1, is

$$W(Z) = \int_0^Z Yf(Y)dY + Z\{F[\theta_{AB}(Z)] - F(Z)\} + (1-\gamma)\int_{\theta_{AB}(Z)}^{\theta_{BC}} Yf(Y)dY + \int_{\theta_{BC}}^{\infty} Yf(Y)dY + \int_{\theta_{BC}}^{\theta_{CL}} \{(\delta-1)\lambda TY - K - \delta N\} f(Y)dY + \int_{\theta_{CL}}^{\infty} \{(\delta-1)TY - K - \delta N\} f(Y)dY \quad . \tag{3.1}$$

This is messy, but straightforward. The first four terms give the output of small non-adjusters (type *M*), adjusters (type *A*), bounders (type *B*), and (combined) concealers and fully compliant taxpayers (types *C* and *L*). The fifth and sixth give the social gain from the transfer from individuals to government of tax payments, net of administration, and compliance costs. Note that the cost incurred by concealers, $\Gamma(Y)Y$, is treated as a transfer, and so does not appear in *W*; as will be seen, this is inessential to the qualitative results.

A useful benchmark for thinking about the optimal threshold in this context is provided by the analysis in Keen and Mintz (2004) of the choice of *Z* when compliance is perfect and there are no behavioral responses. In this case, the social cost of a marginal increase in the threshold is the social value of the revenue foregone, $\delta TZf(Z)$. The marginal social benefit, on the other hand, is the saving in administrative cost to the government and the saving to the private sector in compliance costs and tax paid, $(\delta N + K + TZ)f(Z)$. Equating the two, the optimal threshold in this simple setting is given by

$$Z^{KM} = \frac{\delta N + K}{(\delta - 1)T} \quad . \tag{3.2}$$

This convenient closed form, notably independent of the distribution of taxable income, has proved simple enough to provide practical guidance for policy.[19] It also serves as a useful benchmark for thinking about the optimal threshold in the richer (and more realistic) environment considered here, in which taxpayers have available to them a variety of behavioral responses. The issue, most simply put—and with relevance to

[18] This is certainly a simplistic view of administrative activity, which would, for example, appropriately also involve verifying claims of income below the threshold. But it captures a central cost concern of any tax administration. (Goyette (2012) provides an intriguing analysis of how the VAT threshold is implemented in Uganda, finding it to be focused more on employment than the turnover test by which the requirement to register is formally defined).

[19] Ebrill et al (2001) discuss and illustrate the applicability of this result. (In its original form, focused directly on the VAT, this also includes in the denominator the ratio of value added to sales, since value added is the effective base when a firm whose suppliers are subject to VAT is itself brought into the VAT).

practical advice in this area—is whether these argue for a higher or a lower threshold than the Keen-Mintz threshold Z^{KM}.

Returning to the problem of maximizing (3.1), one sharp result emerges quickly. Recall first from (2.3) and (2.5) that θ_{BC} and θ_{CL} are independent of *Z*. Then, as discussed after Figure 1, the only effect of a small increase in the threshold is to increase the number of adjusters and reduce the number of bounders by exactly the same amount. Private output, however, does change. Differentiating (3.1) gives

$$W'(Z) = Zf(Z) + F(\theta_{AB}) - F(Z) + Z\{f(\theta_{AB})\theta'_{AB} - f(Z)\} - (1-\gamma)\theta'_{AB}\theta_{AB}f(\theta_{AB})$$

$$= F(\theta_{AB}) - F(Z) > 0 \quad , \tag{3.3}$$

the second equality following on cancelling terms and noting from (2.2) that $\theta'_{AB} = 1/(1-\gamma)$, and the inequality from $\theta_{AB} > Z$. Starting from the situation of Proposition 1, it is thus optimal to raise the threshold until θ_{AB} rises above θ_{BC}, and a different partition applies. Thus

PROPOSITION 2: *It is optimal to set the threshold sufficiently high to eliminate bounding*.

The intuition for this strong result—which, unlike others to come, requires no assumption on the shape of the distribution of income—is straightforward. A unit increase in the threshold allows the $F(\theta_{AB}) - F(Z)$ taxpayers bunched just below it to each produce one unit more. It also induces those just on the margin between bounding and adjusting to switch from the former (producing $(1-\gamma)\theta_{AB}$) to the latter (producing *Z*); but their initial indifference between the two means that this has a negligible impact on their output. Thus only the unambiguous increase in the output of the adjusters remains, making it optimal to eliminate bounders entirely. Of course, the starkness of this result reflects the simplicity of the underlying structure: one might in practice expect some degree of heterogeneity, not modeled here, to result in differing forms of behavior within each income. Nonetheless, in so far as the partitioning in Proposition 1 seems to have an element of truth in describing reality, so too will this implication for optimal policy.

Beyond its technical content, Proposition 2 has lessons for the discourse on informality. It emphasizes that the composition of informality matters for policy, since the beneficial elimination of bounders through the increased threshold is simply converting them into adjusters, with no change in the number of informal operators. And that matters, of course, because there is an impact on aggregate output, which stresses another point: reforms that increase informality, in the sense of leading to higher output in the informal sector, may well be desirable.

Building on Proposition 2, the elimination of bounders simplifies the optimization problem. But it helps to approach what remains in stages.

For this, suppose first that there is no concealing, so that—bounding having been optimally removed—the focus is on how best to respond to the possibility of adjusting.[20] All those with maximum potential income below *Z* will simply report that income. Those with higher potential income compare their net income if they earn and declare their maximum income, $(1-T)Y-K$ to that if they adjust to (just below) the threshold, and so will adjust if and only if their potential income is less than

$$\theta_{AL}(Z) \equiv \frac{K+Z}{1-T} > Z, \tag{3.4}$$

which it will prove helpful to think of as an 'effective' threshold: $\theta_{AL}(Z)$ is the lowest income on which (in the circumstances now being assumed) any tax is actually paid. Social welfare is in this case simply

$$W(Z) = \int_0^{\infty} Yf(Y)dY - \int_Z^{\theta_{AL}(Z)} (Y-Z)f(Y)dY$$

$$+(\delta-1)T\int_{\theta_{AL}(Z)}^{\infty} (Y-Z^{KM})f(Y)dY \tag{3.5}$$

using having been made of (3.2) to facilitate comparison with the *KM* threshold. The first order condition on the choice of *Z* then gives:

PROPOSITION 3: *In the absence of concealing, at a local optimum:* [21]

$$W'(Z) = -(\delta-1)T\theta'_{AL}(Z)(\theta_{AL} - Z^{KM})f(\theta_{AL})$$

$$-(\theta_{AL}-Z)\theta'_{AL}(Z)f(\theta_{AL}) + \left[F\left(\theta_{AL}(Z)\right) - F(Z)\right] = 0 \tag{3.6}$$

where, from (3.4),

$$\theta'_{AL}(Z) = \frac{1}{1-T} \tag{3.7}$$

A first lesson from Proposition 3 is that it is now the effective threshold, θ_{AL}, that is most readily related to the benchmark *KM* threshold, not the statutory threshold *Z* itself (the latter being inferred from the former, from (3.4))—which is a natural perspective to take, since it is around the effective threshold that behavioral responses occur. A second is

[20] This, it should be noted, is a simplified version of the most general problem considered in Keen and Mintz (2004), which goes further in taking account of taxed inputs and allowing output to depend on a labor input to which some disutility is attached.

[21] Using $\theta''(Z) = 0$ and simplifying, the second order condition is that

$$W''(Z) = f(\theta)\theta'(1-\theta') - f(Z) - f(\theta)(\theta')^2 - (\theta-Z)f'(\theta)\theta' < 0,$$

for which it is sufficient that $f'(\theta) > 0$.

that, in contrast to the benchmark *KM* case, the optimal threshold is now sensitive to the underlying distribution of firm size.

To see more precisely the intuition for the characterization in Proposition 3, suppose that the effective threshold is initially set at the *KM* level. The first term on the right of (3.6) then vanishes: by exactly the same argument as underlying (3.4), the welfare impacts on revenue and on costs of compliance and administration of slightly raising the threshold net out to zero. But now there are also effects on real output. A marginal increase in the statutory threshold raises the effective threshold by (omitting the subscripts) $\theta'(Z)$, and the $\theta'(Z)f(\theta)$ individuals affected by this each cut their output by $\theta - Z$; this is a source of loss captured by the second term on the right of (3.6). Another effect of the higher threshold, however, is that it enables the $F(\theta) - F(Z)$ individuals just below the threshold to each produce one unit more; which is a source of social gain captured in the third term on the right. Whether the effective threshold should optimally be set above or below Z^{KM} thus depends on the balance between these two effects.

This in turn is closely related to the concavity of convexity of the distribution function $F(Y)$, as can be seen by using (3.7) to rewrite (3.6) as

$$\theta_{AL} - Z^{KM} = \left(\frac{1-T}{\delta - 1}\right) G(\theta_{AL}, Z) \; - \left(\frac{T}{\delta - 1}\right) f(\theta_{AL})(\theta_{AL} - Z) \tag{3.8}$$

where

$$G(\theta, Z) \equiv \left(\frac{1}{f(\theta)}\right)\big(F(\theta) - F(Z) - (\theta - Z)f(\theta)\big) \quad . \tag{3.9}$$

Hence, recalling from (3.4) that $\theta_{AL} > Z$ and confining attention, for brevity, to the case in which *F* is everywhere either concave or convex:

PROPOSITION 4: *In the absence of concealing, it is necessary but not sufficient (respectively, sufficient but not necessary) for the optimal statutory threshold to be greater (less) than Z^{KM} at a local optimum that F be strictly concave (weakly convex).*

Whether the optimal response to the possibility of adjusting in itself points to a lower or a higher threshold than in the benchmark case—and so to more or less informality—is thus unclear without further restrictions on the size distribution. The most plausible case is that in which the density is decreasing over the relevant range, so that *F* is concave. Responding to adjustment may then mean setting a threshold higher than would otherwise be the case. [22]

Returning to the general case in which there may be concealing as well as adjusting (bounders having been optimally eliminated), welfare can be written as

[22] The simulations in Keen and Mintz (2004), for the more complex of their settings noted above, indeed find thresholds optimally higher than Z^{KM}; and not only for concave *F* but also for uniform (which in the present framework implies an optimal threshold unambiguously lower).

$$W(Z) = \int_0^{\infty} Yf(Y)dY - \int_Z^{\theta_{AC}(Z)} (Y - Z)f(Y)dY + (\delta - 1)T\lambda \int_{\theta_{AC}(Z)}^{\theta_{CL}} \left(Y - \frac{Z^{KM}}{\lambda}\right) f(Y)dY + (\delta - 1)T \int_{\theta_{AC}(Z)}^{\theta_{CL}} (Y - Z^{KM})f(Y)dY \quad (3.10)$$

where $\theta_{AC}(Z)$ is the cut-off income between adjusting and concealing, implicitly defined by

$$(1 - \lambda T - \Gamma(\theta_{AC}))\theta_{AC} - K = Z\,. \quad (3.11)$$

Noting that θ_{CL} is independent of *Z*, differentiating in (3.10) gives

PROPOSITION 5: *At a local optimum,*

$$(\delta - 1)T\theta'_{AC}f(\theta_{AC})\left(\theta_{AC} - \frac{Z^{KM}}{\lambda}\right) = -\theta'_{AC}(\theta_{AC} - Z)f(\theta_{AC}) + [F(\theta_{AC}) - F(Z)]\,. \quad (3.12)$$

where

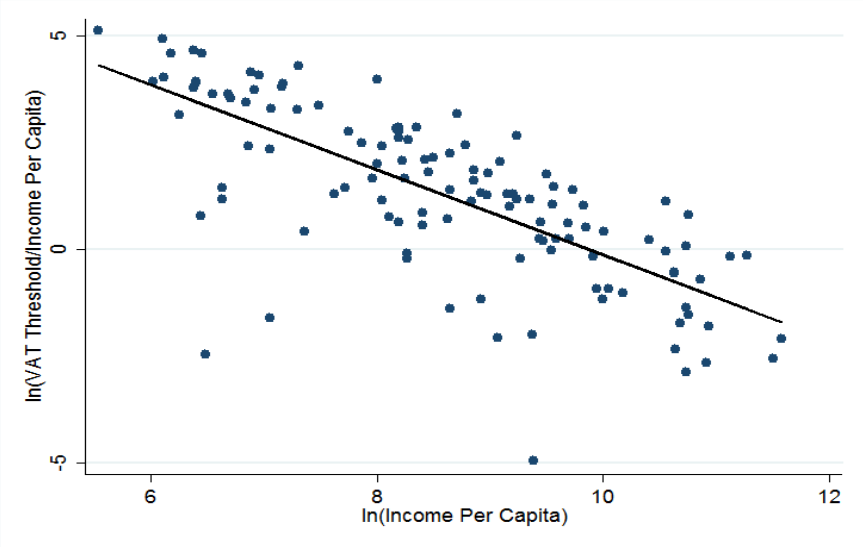
$$\theta'_{AC} = \frac{1}{1 - \lambda T\theta_{AC} - \Gamma(\theta_{AC}) - (\lambda T + \Gamma'(\theta_{AC})\theta_{AC})} > 1 \quad (3.13)$$

The characterization of the optimal threshold in this most general case is evidently similar to that in Proposition 4. The two terms on the right of (3.12) are analogous to the final two on the right of (3.6) and again reflect the shaping of the threshold to address the possibility of adjusting (the difference being that now those on the margin of adjusting are the concealers, whereas in Proposition 4 they were type *L*'s). More fundamentally, on the left of (3.12) the effective threshold is now being compared not directly to the benchmark threshold Z^{KM} but to Z^{KM}/λ; which, since $\lambda < 1$, is strictly larger. In this sense (and independent of the distribution of firms size), responding optimally to concealers calls for setting the threshold higher than would otherwise be the case. Indeed it can be shown, in similar spirit, that the optimal statutory threshold is decreasing in λ;[23] is lower, that is, the greater the proportion of income that is declared. This is readily explained. Returning to the logic underlying the benchmark *KM* result in (3.2), when individuals declare only a proportion their income, the revenue impact is as if they fully declared but paid at a rate of λT. Applying that same logic then implies a threshold of Z^{KM}/λ: the threshold is optimally set higher simply because the revenue loss from raising it is less than it would be if all those paying tax were fully compliant.

[23] The proof, which is cumbersome, is omitted.

It is clear from these results that setting an optimal threshold becomes much more complex when account is taken of the various forms of compliance and non-compliance that are the commonplace of tax implementation. It requires balancing considerations whose relative force likely varies widely both across countries and over time. Where taxpayers are broadly honest, whether by inclination or by features of design and administration, the main additional concern is with the possibility of some avoiding tax by legitimately adjusting below the threshold. The implications of this for the level of the threshold are sensitive to the shape of size distribution of firms, but may plausibly point to a higher threshold than otherwise. Where, however, evasion is the dominant concern—whether in the form of bounders or concealment—the implication is somewhat more clear-cut, the analysis here pointing to a higher threshold than would otherwise be optimal. To the extent that evasion can be expected to be a greater concern in lower income economies, with less well-developed revenue administrations, this suggests that, all else equal, thresholds should optimally be higher there. Of course all else (including the distribution of firm size) is not equal. It is of some interest, nonetheless, that Figure 2, which plots VAT thresholds relative to per capita income against per capita income (both in logs) suggests that this is indeed quite broadly the case in practice

Figure 2. VAT thresholds and per capita income, 2012



Note: Author's calculations using VAT thresholds from IMF compilation and per capita income from the *World Economic Outlook.*

IV. Multiple Obligations and Thresholds

In practice, individuals and firms commonly face not just one potential tax (or other obligation), as assumed so far, but several—each with its associated threshold.[24] These may relate to different dimensions of behavior: labor law requirements, for instance, are often defined in terms of number of employees, while tax obligations are arise when income or turnover exceed some limits. Or they may relate to the same dimension: tax systems may, for example, use threshold levels of turnover not only to define the obligation to register for the VAT but also to mark the transition between various forms of income tax (some form of presumptive tax for the smallest, in many cases, with transition to full accrual-based taxation for the largest).[25]

While notches, kinks and thresholds have received considerable attention in recent years, the interplay between the many such features that are faced in reality has not. They can clearly greatly complicate individuals' and firms' decisions: choosing to bunch below the notch for one obligation, for example, will generally have implications for liability under others. An Indian manufacturing firm that chooses to hire more than 9 employees, for instance, and so become subject to the Factories Act, may consequently find it less attractive to maintain its turnover below the 10,000 rupees at which it is required to charge Central Excise Duty. This in turn makes policy makers' design problem in setting thresholds still harder—and, as in this example, may call for coordination between agencies that have quite different mandates but whose decisions affect the same individuals and firms.

One very practical issue that arises is whether there is merit in setting the threshold for different obligations or treatment at the same level, as sometimes seems to be presumed. One strategy recommended for the treatment of small businesses, for instance, has been to set a common turnover threshold for both the VAT and some form of self-assessment under the income tax (International Tax Dialogue, 2007).

To begin exploring some of the implications of multiple obligations and thresholds, suppose now that firms face two distinct taxes, each with the same structure as above, indicated by subscripts 1 and 2. These may differ in all respects: tax rate, threshold and implementation costs. (But they need not: indeed the case in which they are identical is an important benchmark, in that if each were then set in isolation of the other, they would indeed be given the same threshold—which poses crisply the question of whether that is indeed optimal when the interactions between them are recognized). These two taxes might be thought of, for instance, as the VAT and an income tax, with an assumption of a fixed relationship between income and turnover enabling the two to be expressed in the same common unit of sales. Less literally—but as a step away from the blinkered perspective of looking only at some subset of the many obligations firms may face—the framework might also be thought of as reflecting tax and non-tax obligations (such as

[24] Note that the concern here is with multiple obligations, not with a single obligation that may take different forms (such as quarterly versus annual reporting for the VAT).

[25] See for instance the description of small business tax regimes in Inter-American Development Bank (2013)

factory act requirements), with the private and social costs and benefits of the latter translated into tax-equivalent terms.

Leaving a general treatment to future work, we shall simply suppose that all firms are intrinsically honest: there are no bounders or concealers, so the only behavioral concern is adjusting. Without loss of generality, we also assume that $Z_1 \leq Z_2$; so the only options that the taxpayer has are to pay both taxes, pay only 1, or pay neither. Broadly two types of partition can then arise.

The first arises when Z_2 is sufficiently close to Z_1 that $K_1 + T_1 Z_2 > Z_2 - Z_1 > 0$. Then

$$(1 - T_1)Z_2 - K_1 \leq Z_1 \tag{4.1}$$

and all those declaring under the threshold for 2 but above that for 1 would find it still better to declare under the threshold for 1: intuitively, the threshold for 2 is so low as to be in the 'hole' just above the threshold for 1. The only relevant choice is then that between paying both taxes and paying none. Using an absence of subscripts to indicate addition over both taxes (so that $T \equiv T_1 + T_2$, and so on), the critical level of income below (above) which it is privately best to pay neither tax (both taxes) is then given by[26]

$$\theta \equiv \frac{K + Z_1}{1 - T} \tag{4.2}$$

The outcome is thus as in the top panel of Figure 3: a single, large mass of firms just below the lower statutory threshold, above which there is a single hole extending above the higher threshold, with neither tax being paid until an effective threshold is reached, above both thresholds, at which point both are paid. We refer to this as a Type I partition.

For the second possibility, suppose now that the thresholds satisfy the converse of (4.1), so that Z_2 lies outside the hole associated with tax 1. Among those with potential income below Z_2, those with income less than

$$\theta_1 \equiv \frac{K_1 + Z_1}{1 - T_1} \tag{4.3}$$

will adjust to below Z_1; the rest will pay tax 1.[27] The fundamental difference from the previous case, however, is that for those above Z_2 it will be better to adjust out of tax 2 only than to adjust out of both. More precisely, they will choose to adjust out of tax 2 (only) rather than pay both if and only if their income is below the level θ_2 at which

$$(1 - T)\theta_2 - K = (1 - T_1)Z_2 - K_1 \; ; \tag{4.4}$$

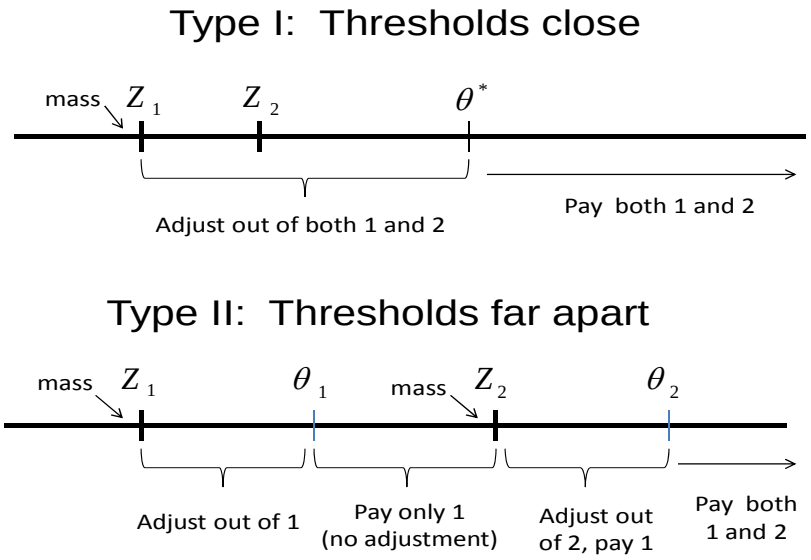
[26] That $\theta > Z_2$, as drawn (and used in proving Proposition 6 below) follows from (4.2) and (4.1).
[27] That $\theta_1 < Z_2$ follows from (4.3) and the negation of (4.2).

that is, below[28]

$$\theta_2 = \frac{(1-T_1)Z_2 + K_2}{1-T} \quad . \tag{4.5}$$

The second possibility is thus as in the lower panel of Figure 3: now there is a mass of firms just below, and a hole above, each threshold. We call this a Type II partition.

Figure 3: Partitions with two obligations



The question then arises as to whether these very different forms of partition can be welfare-ranked. Is it better to set the thresholds for distinct obligations at the same or similar levels, generating a large mass below the lower of the two (Type I), or to set them far enough apart to induce some mass below both thresholds (Type II)? The following establishes that the latter is preferred under quite a weak condition:

PROPOSITION 6: *For given Z_1, if $T_1 Z_1 > N_1$ then it is optimal to set Z_2 high enough to induce a partition of type II.*

Proof: See appendix.

This finding that a marked dissimilarity of thresholds is optimal—even, it should be noted, if all aspects of the two instruments are identical—points to the importance of the interplay between distinct obligations. To see the logic, suppose that the thresholds of a VAT and a Factory Act are calibrated to apply at roughly the same levels of activity, so that initially firms comply with either both or neither, with a large mass below the common threshold. Then by raising the threshold for the Factory Act sufficiently (from,

[28] That $\theta_2 > Z_2$ follows from (4.5) and $T_2 > 0$.

say, 10 workers to 20) some firms will be induced to so increase their scale (from 9, say, to 19) that while continuing to remain out of the Factory Act it becomes worth their while to become subject to VAT. In this way, raising the threshold for one obligation eventually induces a mass of firms into compliance with the *other* obligation; thus it is that the condition in Proposition 6 for the significantly higher threshold for tax 2 to be optimal turns on the net tax receipts from the other tax, 1.

A different conclusion might be reached, of course, if there were some economies of scope in complying with or administering the different obligations; and there is of course some merit of simplicity in having fewer thresholds rather than more. What emerges clearly, however, is the importance of recognizing the multiplicity of obligations to which firms are generally subject, and the interplay between them. At one level, this raises yet more issues that the simple language of informality can hardly cope with: in the example above, for instance, it may well be that, even at an optimum, some firms are 'formal' with respect to obligation 1 but 'informal' in relation to obligation 2. More fundamentally, it points to a treatment of partition issues that looks holistically at the full set of obligations placed on firms. This perspective appears to be as absent from policy making as it has been from the formal literature. And it opens up further lines for enquiry. It might be, for instance, that each agency, in setting thresholds related to its own sphere of interest, neglects the impact on aspects of compliance of interest to other agencies, with consequent inefficiencies that a more coherent approach could avoid.

V. Concluding

The simple models of this paper, and the characterizations and propositions to which they gives rise, highlight a number of important issues in the current literature on taxation and informality.

First, the results above make clear that the term 'informal' as used in the literature is imprecise and can be very misleading. What the models reveal, and what is surely true in practice, is a range of compliant and non-compliant behaviors which are often and unthinkingly lumped together as a single entity under the label of 'informality.' As we have shown, a key policy instrument like the tax threshold can have an impact on social welfare primarily through intricate effects on the composition of informality rather than through its overall level. Among the results is one with a practical implication: where concerns on evasion are stronger, as for example in many developing countries, there is an argument for tax thresholds to be set higher than they would be in the absence of behavioral response.

Second, the paper has highlighted an issue which has not received the attention it surely deserves. When firms and individuals face multiple forms of tax and non-tax obligations with different thresholds, quite complex patterns of compliance, adjustment and evasion can be generated. Specifically, firms can choose to adjust out of some obligations but not others. How should thresholds be set in such a setting? The simple analysis of this paper, with two obligations and two thresholds, points to quite a rich set of possibilities.

For example increasing the threshold for one tax obligation may induce firms, even if they continue to remain below that threshold, to cease adjusting out of the other--which may give rise to social gain. A rich research agenda awaits, including tackling issues of coordination across different agencies which manage the different tax and non-tax obligations.

The analysis here has many limitations, of course, notably in the somewhat mechanical modeling of opportunities to evade tax, which ignores, not least, the element of risk that has featured prominently in much of the literature[29] The treatment has also been very partial, in the sense that it has focused on just one policy instrument—the threshold—when the analysis itself makes clear than many others, including tax structure and enforcement strategies, have a key role in shaping the kinds of partitions with which we have been concerned. It has served, nonetheless, to illustrate the potential value of the approach to issues of non-compliance and informality described at the outset, moving away from broad-brush notions of informality and towards understanding how policy choices induce partitions of taxpayers by the nature of their non-compliance (or compliance).

This focus, it should be noted, resonates very strongly with current trends in the practicalities of revenue administration. These are often built around notions of 'taxpayer segmentation,' implemented through organizational structures that enable distinct treatment of groups presenting different compliance risks. And these, in turn, are often based on a simple partition by size, in many cases beginning with large taxpayer units and with an intention to develop subsequently medium- and small-taxpayer services. The broad parallel with the partition linking size to compliance behavior established here is striking, so that the present analysis provides some underpinning for such an administrative strategy. It may also point the way towards richer models that can help administrations identify, in their particular circumstances what, and where, the main compliance risks are likely to be. Perhaps most important, however, the approach here stresses the need for a conscious simultaneity in policy design and practical segmentation strategies, with the former, through the choice of threshold and other instruments, shaping the compliance risks that the latter are intended to address.

[29]As a way of thinking about the VAT, the model also neglects the links between firms that recent work suggests can be important in propagating or curtailing forms of non-compliance under the VAT (De Paula and Scheinkman (2010), Pomeranz (2011)).

Appendix: Proof of Proposition 6

Welfare in a type I partition is

$$W^A(Z_1) = \int_0^{Z_1} Yf(Y)dy + Z_1\{F(\theta) - F(Z_1)\} + \int_\theta^\infty Yf(Y)dY + \int_\theta^\infty \{(\delta - 1)TY - \delta N - K\}f(Y)dY, \quad \text{(A. 1)}$$

which, note, is independent of Z_2. In a type II partition, which, recalling (4.1), arises with a threshold of

$$Z_2(\varepsilon) = \frac{K_1 + Z_1}{1 - T_1} + \varepsilon \quad \equiv Z_2^* + \varepsilon \quad \text{(A. 2)}$$

for any $\varepsilon > 0$, welfare is

$$\begin{aligned} W^B(Z_1, Z_2(\varepsilon)) &= \int_0^{Z_1} Yf(Y)dY + Z_1\{F(\theta_1) - F(Z_1)\} \\ &+ \int_{\theta_1}^{Z_2} Yf(Y)dy \\ &+ Z_2\{F(\theta_2) - F(Z_2)\} + \int_{\theta_2}^\infty Yf(Y)dy + \int_{\theta_1}^{Z_2} \{(\delta - 1)T_1Y - \delta N_1 - K_1\}f(Y)dY \\ &\int_{\theta_2}^\infty \{(\delta - 1)TY - \delta N - K\}f(Y)dY + \{(\delta - 1)T_1Z_2 - \delta N_1 - K_1\}\{F(\theta_1 - \theta_2)\}. \quad \text{(A. 3)} \end{aligned}$$

Noting that $\lim_{\varepsilon\to 0} \theta_1 = \theta$ (from (A.2) and (4.3)) and $\lim_{\varepsilon\to 0} \theta_2 = Z_2$ (from (A.2) and (4.5)), subtracting (A.3) from (A.1) and taking limits gives, canceling and collecting terms and using (A.2),

$$\lim_{\varepsilon\to 0} W^B - W^A = \lim_{\varepsilon\to 0} \delta(T_1Z_2^* - N_1)\{F(\theta) - F(Z_2^*)\}\ , \quad \text{(A. 4)}$$

from which, since $\theta > Z_2^*$, as noted in the text, the result follows.

References

Alm, James, Jorge Martinez-Vazquez, and Sally Wallace (Eds.), 2004, *Taxing the Hard-to-Tax* (Amsterdam: Elsevier).

Almunia, Miguel and David López-Rodríguez, 2013 "Firms' responses to tax enforcement policies: Evidence from Spain" Mimeo, University of Warwick.

Auriol, Emmanuelle, and Michael Warlters, 2005, "Taxation base in developing countries," *Journal of Public Economic*s, Vol. 89, pp. 625–646.

Bastani, Spencer and Hakan Selin, 2012, "Bunching and non-bunching at kink-points of the Swedish tax schedule," CESifo Working Paper No. 3865.

Burgess, Robin and Nicholas H Stern, 1993, "Taxation and development, "*Journal of Economic Literature*, Vol. 31, pp. 762-830.

Chatterjee, Urmila and Philippe Wingender. 2011. "Tax-Threshold and bunching: Evidence from Indian firms." Mimeo, IMF.

Chatterjee, Urmila and Ravi Kanbur, 2013. "Regulation and non-compliance: Magnitudes and patterns for India's Factories Act." World Bank Policy Research Working Paper, No. 6755.

Coolidge, Jacqueline, 2012, "Findings on tax compliance cost surveys in developing countries," *eJournal of Tax Research*, Vol. 10, pp. 250-287.

Dabla-Norris, Era, Mark Gradstein and Gabriela Inchauste, 2008, "What causes firms to hide output? The determinants of informality," *Journal of Development Economics*, Vol. 85, pp. 1-27.

De Paula, Áureo and José A. Scheinkman, 2010, "Value-added taxes, chain effects and informality," *American Economic Journal: Macroeconomics*, Vol. 2, pp. 195-221.

Dharmapala, Dhammika, Joel Slemrod and John Douglas Wilson, 2011, " Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs," *Journal of Public Economics*, Vol. 95, pp. 1036-1047.

Ebrill, Liam, Michael Keen, Victoria Summers and Jean Paul Bodin, 2001, *The Modern VAT* (Washington DC: International Monetary Fund).

Erard, Brian and Chin-Chin Ho, 2001, "Searching for ghosts: Who are the nonfilers and how much tax do they owe?" *Journal of Public Economics*, Vol. 81, pp. 25–50.

European Commission, 2004, "European tax survey," Directorate General Customs Union and Taxation, Working Paper No.3.

Garicano, Luis, Claire Lelarge and John van Reenen, 2013, "Firm size distortions and the productivity distribution: Evidence from France," Mimeo, London School of Economics.

Gauthier, Bernard and Mark Gersovitz, 1997, "Revenue erosion and evasion in Cameroon, 1993," *Journal of Public Economics*, Vol. 64, pp. 407–424.

Gauthier, Bernard and Ritva Reinikka, 2001, "Shifting tax burdens through exemptions and evasion: An empirical Investigation of Uganda." Mimeo, World Bank.

Goyette, Jonathan, 3012, "Optimal VAT threshold: Official vs. effective enforcement." Mimeo, University of Sherbrooke.

Hindriks, Jean, Michael Keen and Abhinay Muthoo, 1999, "Corruption, extortion and evasion," *Journal of Public Economics*, Vol. 74, pp. 395-430.

Inter-American Development Bank, 2013, *More than Revenue: Taxation as a Development Tool* (Washington DC: Inter-American Development Bank).

International Tax Dialogue, 2007, "Taxation of small and medium enterprises," available at http://www.itdweb.org/smeconference/conferencepapers.aspx

Kanbur, Ravi, 2009, "Conceptualising informality: Regulation and enforcement," *Indian Journal of Labour Economics*, Vol. 52, No. 1, pp. 33-42.

Keen, Michael, 2013, "Taxation and development—Again," pp. 13-41 in Clemens Fuest and George Zodrow (Eds.), *Studies of Critical Issues in Taxation and Development* (Cambridge: MIT Press).

Keen, Michael and Jack Mintz, 2004, "The optimal threshold for a value-added tax," *Journal of Public Economics*, Vol. 88, pp. 559-576.

Keen, Michael and Alejandro Simone, 2004, "Tax policy in developing countries: Some lessons from the 1990s, and some challenges ahead," pp. 302-352 in *Helping Countries Develop: The Role of Fiscal Policy*, edited by Sanjeev Gupta, Ben Clements, and Gabriela Inchauste (Washington, DC: International Monetary Fund).

Kleven, Henrik J. and Mazhar Waseem, 2013, "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan," *Quarterly Journal of Economics*, Vol. 128, pp. 669-723.

Mubiru, Alex, 2010, "Committee of Ten policy brief: Domestic resource mobilization across Africa: Trends, challenges and policy options," available at *www.afdb.org/.../C-10%20Note%202%20English%20(final)_for%20pos...*

Onji, Kazuki, 2009, "The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax," *Journal of Public Economics*, Vol. 93, pp. 766-775.

Organisation for Economic Co-operation and Development, 2008, "Governance, Taxation and Accountability: Issues and Practices," Development Assistance Committee Guidelines and Reference Series (Paris: OECD).

Organization for Economic Cooperation and Development, 2009, *Is Informal Normal?* (Paris: OECD).

Pomeranz, Dina, 2011, "No taxation without information: Deterrence and self-enforcement in the value added tax," forthcoming in *Tax Administration Review*.

Saez, Emmanuel, 2010, "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, Vol. 2, pp. 180-212.

Slemrod, Joel, 2001, "A general model of the behavioral response to taxation," *International Tax and Public Finance* Vol. 8, pp. 119-128.

Slemrod, Joel, 2010, "Buenas noches: Lines and notches in tax system design." Mimeo, University of Michigan.

Tybout, James R., 2000, "Manufacturing firms in developing countries: How well do they do, and why?" *Journal of Economic Literature*, Vol. 38, pp. 11–44.

Zee, Howell, 2005, "Notes and communications," *De Economist*, Vol. 153, pp. 461-471.

OTHER A.E.M. WORKING PAPERS

WP No	Title	Fee (if applicable)	Author(s)
2014-10	On the Political Economy of Guest Workers Programs in Agriculture		Richard, B.
2014-09	Groupings and the Gains from Tagging		Kanbur, R. and M. Tuomala
2014-08	Economic Policy in South Africa: Past, Present, and Future		Bhorat, H., Hirsch, A., Kanbur, R. and M. Ncube
2014-07	The Economics of China: Successes and Challenges		Fan, S., Kanbur, R., Wei, S. and X. Zhang
2014-06	Mindsets, Trends and the Informal Economy		Kanbur, R.
2014-05	Regulation and Non-Compliance: Magnitudes and Patterns for India's Factories Act		Chatterjee, U. and R. Kanbur
2014-04	Urbanization and Agglomeration Benefits: Gender Differentiated Impacts on Enterprise Creation in India's Information Sector		Ghani, E., Kanbur, R. and S. O'Connell
2014-03	Globalization and Inequality		Kanbur, R.
2014-02	Should Mineral Revenues be Used for Countercyclical Macroeconomic Policy in Kazakhstan?		Kyle, S.
2014-01	Performance of Thailand Banks after the 1997 East Asian Financial Crisis		Mahathanaseth, I. and L. Tauer
2013-20	What is a "Meal"? Comparing Methods to Determine Cooking Events		Harrell, S., Beltramo, T., Levine, D., Blalock, G. and A. Simons
2013-19	University Licensing of Patents for Varietal Innovations in Agriculture		Rickard, B., Richards, T. and J. Yan
2013-18	How Important Was Marxism for the Development of Mozambique and Angola?		Kyle, S.
2013-17	Social Protection, Poverty and the Post-2015 Agenda		Fiszbein, A., Kanbur, R. and R. Yemtsov

Paper copies are being replaced by electronic Portable Document Files (PDFs). To request PDFs of AEM publications, write to (be sure to include your e-mail address): Publications, Department of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853-7801. If a fee is indicated, please include a check or money order made payable to Cornell University for the amount of your purchase. Visit our Web site (*http://dyson.cornell.edu/research/wp.php)* for a more complete list of recent bulletins.