NBER WORKING PAPER SERIES

# WHY CAN MODERN GOVERNMENTS TAX SO MUCH? AN AGENCY MODEL OF FIRMS AS FISCAL INTERMEDIARIES

Henrik Jacobsen Kleven
Claus Thustrup Kreiner
Emmanuel Saez

Working Paper 15218
http://www.nber.org/papers/w15218

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2009

We would like to thank David Albouy, Tim Besley, Saki Bigio, Raj Chetty, Alain Jousten, Wojciech Kopczuk, Camille Landais, Thomas Piketty, Monica Singhal, Joel Slemrod, and numerous seminar participants for helpful discussions and comments. This paper is an output of the International Growth Centre (IGC). Additionally, financial support from NSF Grant SES-0850631 is gratefully acknowledged. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries
Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez
NBER Working Paper No. 15218
August 2009
JEL No. H11,H20


## ABSTRACT

This paper presents a simple agency model to explain why third-party income reporting by employers dramatically improves income tax enforcement. Modern firms have a large number of employees and carry out complex production tasks, which requires the use of accurate business records. Because such records are widely used within the firm, any single employee can denounce collusive tax cheating between employees and the employer by revealing the true records to the government. We show that, if a firm is large enough, such whistleblowing threats will make tax enforcement successful even with low penalties and low audit rates. Embedding this agency model into the standard Allingham-Sandmo tax evasion model, we show that third-party reporting improves tax enforcement if the government disallows self-reported losses or audits such losses more stringently, which fits with actual tax policy practices. We also embed the agency model into a simple macroeconomic growth model where the size of firms grows with exogenous technological progress. In early stages of development, firms are small, tax rates are severely constrained by enforcement, and the size of government is too small. As firm size increases, the enforcement constraint is slackened, and government size is growing. In late stages of development, firm size is sufficiently large to make third-party tax enforcement completely effective and government size is socially optimal.



Henrik Jacobsen Kleven
Department of Economics & STICERD
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
H.J.kleven@lse.ac.uk

Emmanuel Saez
Department of Economics
University of California, Berkeley
549 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

Claus Thustrup Kreiner
Institute of Economics
University of Copenhagen
Studiestraede 6
DK-1455 Copenhagen K
Denmark
claus.thustrup.kreiner@econ.ku.dk

# 1 Introduction

The size of governments has expanded dramatically over the 20th century. A central element of this expansion has been the ability of governments to extract a substantial fraction of national products through taxation without destroying economic growth. In all advanced economies, most taxes are collected through third-party institutions such as private or public employers, banks, investment funds, and pension funds. These entities (which we call "firms") generally have a large number of employees, clients, or business partners. Therefore, they need to use accurate and rigorous records to carry out their complex business activities. Firms report taxable income—such as corporate profits, compensation paid to employees, or capital income paid to clients—directly to the government, and therefore act as a third party between households and the government. They also often withhold taxes on behalf of the government so that tax payments take place "as-you-go".[1]

It is widely known in the tax law literature (e.g., Surrey 1958; Lederman 2009) as well as among tax practitioners (e.g., Bird 2003; OECD 2004, 2006) that tax enforcement is excellent whenever such third-party reporting is in place, and that enforcement is weak—even in the most advanced economies—when such third-party reporting is not in place, as in the case of small family businesses. Therefore, as a first approximation, tax enforcement is successful if and only if third-party reporting covers a large fraction of taxable income. For example, the most recent US Tax Compliance Measurement Study (Internal Revenue Service, 2006) shows that individual income tax evasion rates is 53.9% when there is "little or no" information reporting, but that the evasion rate is less than 5% when there is substantial information reporting.[2]

In spite of its central importance, the theoretical literature on tax evasion has not devoted much attention to the issue of third-party reporting or tried to explain why such a system is successful. Indeed, most of the modern literature on tax evasion follows on the seminal study by Allingham and Sandmo (1972), which used the Becker (1968) model of crime and focuses on a situation with no third-party reporting, i.e., on the case where enforcement is never successful

[1]The withholding system is useful to individuals or companies when there are credit constraints, a point we will not investigate in this paper where we focus only on informational aspects.

[2]Similarly, Kleven et al. (2009) study an income tax audit experiment in Denmark and find that, although purely self-reported income constitutes only about 8% of total reported income, it accounts for about 90% of detected evasion. Eurostat (2007) uses a questionnaire on undeclared work in the European Union and shows that it is concentrated primarily among the self-employed providing direct services to households.

in practice and which covers a minor part of taxation in advanced economies.[3] The Allingham-Sandmo model generates a key puzzle: why are compliance rates so high in developed countries given that audit rates and penalties for tax evasion are generally very low?[4]

Our paper sets out a three-tiered agency model to provide a simple micro-foundation for the success of third-party reporting. In the model, the government is the top tier (principal) trying to extract tax revenue from individual income earners (bottom tier agents) who are employed or paid by firms (middle tier). The firm acts as a third party that reports income on behalf of individuals. Although we focus on the case where individuals are employees of the firm, the model can easily be applied to a situation where individuals are clients investing their savings and receiving capital income from a financial institution, or shareholders receiving profits from the firm. When a firm is large and complex, using detailed business records—such as accounting books, details of purchases and sales, or payroll accounts listing individual wages and salaries—is extremely valuable for productivity. Such records are widely used within the firm and hence many employees know about them.

In principle, the firm and its employees could collude to report smaller incomes—salaries and profits—to the government than those actually earned. Under perfect information and commitment between the firm and individuals, there would be no reason for breaking the collusion. In practice, breakdowns can occur because of random shocks such as conflicts between employees and the employer, moral concerns of a newly hired employee, or an employee mistakenly showing the true business records to tax inspectors. Breakdowns can also occur as a result of rational whistleblowing if the government provides rewards to whistleblowers and firms cannot make employees commit not to whistleblow ex-ante. In our model, we assume that each employee has the option of reporting cheating to the government by divulging the true business records to the government. When a firm has many employees, breakdowns of collusion will occur with a high probability. Critically, it is the combination of a large number of informed employees and the existence of business records evidence, which makes third-party tax enforcement successful.[5]

[3]See Andreoni et al. (1998), Cowell (1990), Schneider and Enste (2000), Slemrod and Yitzhaki (2002), Slemrod (2007), and Shaw, Slemrod, and Whiting (2008) for comprehensive surveys.

[4]As Andreoni et al. (1998) conclude at the end of their survey (p. 855): "The most significant discrepancy that has been documented between the standard economic model of compliance and real-world compliance behavior is that the theoretical model greatly over-predicts noncompliance." Various studies suggest that high compliance rates may be explained by psychological or behavioral aspects such as social norms, tax morale, patriotism, guilt and shame (e.g., Cowell 1990, chapter 6; Andreoni et al. 1998, Section 8). In this paper, we propose instead a theory explaining high compliance based on information.

[5]Our model focuses on *internal* information sharing within the firm. However, firms also share information

We embed this agency model into the Allingham and Sandmo (1972) model of tax evasion assuming that some income items (such as wages and salaries) are third-party reported, while other income items are solely self-reported (such as self-employment income). We first demonstrate a surprising third-party irrelevance result: even if the government can observe third-party reported items perfectly at no cost, third-party reporting will be entirely undone by individual tax filers who adjust self-reported income correspondingly. However, this irrelevance result depends critically on two assumptions: self-reported losses are allowed and audit rates are independent of the level (or sign) of self-reported income. In practice, self-reported income losses are often disallowed to count against other income items, and tax audits are concentrated on self-reported income and especially self-reported income losses. In those circumstances, the irrelevance result no longer holds and third-party reporting does reduce overall tax evasion.

The last part of the paper embeds our agency model into a simple macroeconomic growth model where the size and complexity of firms grows with exogenous technological progress. In this model, a representative individual has preferences over private and public goods. In the absence of enforcement problems, taxes are non-distortionary and should be set to finance public goods according to the classical Samuelson rule. We model utility functions such that the public good has an income elasticity equal to one, implying that the first-best tax rate is constant along the path of economic growth. With tax enforcement constraints, however, there are three regimes over the process of development. In the earliest stage, firms are very small and untaxable, and therefore the government raises no tax revenue and supplies no public goods. In the middle stage, firm size is large enough that firms start becoming taxable provided that the tax rate is not too high. In that stage, the enforcement constraint is binding, and the government tax rate and public goods provision are below the first-best level but growing over time. In the latest stage, firms have become so large that, even under the first-best tax rate, firms choose to remain in the formal sector and pay taxes. The government imposes the first-best tax rate and government size relative to output is optimal and stable over time.[6]

This simple macro model can account for the historical growth in government size over the

with *external* parties such as other businesses and individual clients, shareholders, or debt holders. The number of such external parties also grows with economic development, making tax collusion more difficult as in our internal information sharing model.

[6]Although we present the theory in the context of a benevolent government maximizing the welfare of a representative household, the story is consistent with a Leviathan view of government where self-interested politician-bureaucrats maximize tax revenue.

last century and the stability of government size since the 1970s in the richest economies. The theoretical story does not rely on demand for public goods effects or political economy effects. Our theory shows that technological progress and economic growth leads to large and complex firms, which can then be easily taxed. Therefore, our theory shows that capitalism—in the sense of the emergence of large and complex firms using rigorous accounting—is a necessary condition for the rise of large welfare state governments, which fund public programs such as welfare programs, social insurance programs, retirement benefits, and education. This can be seen as a *Marxist theory in minor mode*: rather than leading to revolution and communism, capitalism, by relaxing the tax enforcement constraint, breads large welfare states.

The paper is organized as follows. Section 2 reviews the related literatures. Section 3 presents our micro-model of third-party tax enforcement and then embeds this model into a standard Allingham-Sandmo model of tax evasion. Section 4 embeds the micro-model in a simple macroeconomic framework which accounts for the evolution of government size over the course of development. Finally, Section 5 offers concluding remarks.

# 2 Review of Related Literature

## 2.1 Literature on Tax Evasion and Tax Enforcement

Our agency model of third-party tax enforcement contributes to the large literature on tax evasion and tax enforcement. A few previous studies have incorporated information reporting into tax evasion models. Yaniv (1992) made the simple but important point that, if the employer and employees can collude, then third-party reporting cannot help tax enforcement. Our paper starts from this benchmark and shows that this collusion equilibrium is fragile in the presence of verifiable business records and many employees.

Recently, a number of studies have made important progress in modelling the problem of tax enforcement. First, Gordon and Li (2005) develop a model where the government can collect taxes only from formal firms defined as those connected to the financial sector. Access to credit is indeed one way in which using rigorous accounting books improves productivity. They show how the lessons from optimal tax theory are drastically changed in this environment and fit much better with actual tax policies. Instead of considering a reduced-form model of tax enforcement, our paper zooms in on the micro-foundations of third-party reporting by explicitly modelling the tax evasion game in a three-tiered agency model and is fully consistent with the

contribution of Gordon and Li (2005). Second, Kopczuk and Slemrod (2006) set out a simple model to show how the network of firm-to-firm arm-length transactions can help the government enforce taxation. In that context, the authors demonstrate formally the important point that value-added taxes and retail sales taxes are no longer equivalent as value added taxes are easier to enforce using firm-to-firm transaction information. Third, Keen (2007) shows that a value-added-tax allows to tax informal suppliers because formal businesses cannot take a deduction for purchases from informal suppliers. Our paper focuses primarily on the within-firm information network rather than the across-firm information network and is therefore complementary to the Kopczuk-Slemrod and Keen papers.[7] Finally, a number of studies in the corporate income tax evasion literature have shown that the internal organization or the external activities of firms can affect their tax reporting decisions.[8]

At a broad level, our paper is related to the large theoretical literature on mechanism design and implementation, especially work on mechanism design in environments with complete information among agents (such as employees in a firm). This literature has been surveyed by Moore (1992), who showed that cross-reporting is a powerful instrument that often allows the principal to elicit truthful information from agents, at least in a non-cooperative setting and if large fines are feasible. Our model of third-party reporting encompasses this basic idea, although we allow for the possibility of collusive behavior and assume (realistically) that there is a upper bound on the size of fines, both of which makes tax enforcement harder. In this case, it is not always feasible to achieve truthful reporting and the efficacy of enforcement depends on firm size. We come back to these mechanism design issues below, where we discuss the potential for non-conventional tax enforcement mechanisms to improve the truthfulness of income reporting.

## 2.2 Literature on the Growth of Government

Our macro model contributes to a very long literature trying to explain the growth of government. A number of theories have been put forward. First, the famous "Wagner's law" (after

[7] We discuss briefly how the network of firm-to-firm transactions can also help enforcement as firms can also denounce tax cheating of other firms.

[8] On the internal side, Crocker and Slemrod (2005) develop a shareholder-manager agency model with tax evasion showing that penalties imposed on managers are more effective in reducing evasion than penalties imposed on shareholders. Chen and Chu (2005) show that the evasion decision of the firm's owner affects the optimal compensation scheme offered to employees and hence creates a distortion in the manager's effort and reduces the efficiency of the contract. On the external side, Bayer and Cowell (2005) show that imperfect competition between firms have important consequences for the efficiency effects of corporate tax audits.

the German economist Adolph Wagner, 1835-1917) focuses on the demand side and posits that public goods have an income elasticity above one (see Musgrave, 1966, for a detailed exposition and analysis). Second, Baumol's cost disease theory focuses on the supply side and posits that, over the course of development, productivity in the private sector increases while productivity in the public sector stagnates, leading to a growth of government spending relative to GDP (Baumol and Bowen, 1966; Baumol, 1967). Third, Peacock and Wiseman (1961) proposed a "ratchet effect theory", whereby temporary shocks such as wars raise government expenditures, which do not fall back after the shock as social norms regarding the proper level of public goods and taxation are permanently affected by the temporary shock. Notice that the Wagner, Baumol, and ratchet effect theories cannot explain the long period of stable government expenditures before the 20th century, a period with some economic growth and with many wars creating temporary spending shocks. Fourth, the Leviathan theory posits that governments are controlled by self-interested politician-bureaucrats, unchecked by electoral constraints (Brennan and Buchanan, 1980), and hence maximize revenue under constitutional and fiscal constraints. Although proponents of the Leviathan theory have focused primarily on public choice and constitutional aspects, this theory is entirely consistent with the importance of tax enforcement constraints that we emphasize in this paper. Fifth, a large literature on political economy considers the role of voting, lobbying, corruption, and political constitutions for the size of government. This literature has proposed that the democratization and increased political power of the poor have played an important role for the growth of government (Acemoglu and Robinson, 2000). Moreover, substantial attention has been paid to the relationship between changes in income distribution and voters' demand for redistribution (Peltzman, 1980; Lindert, 2004).

In addition to these hypotheses, a number of studies have pointed out that there are fiscal capacity constraints to government growth (e.g., Kau and Rubin, 1981; Bird, 1989, 1992; Peltzman, 1980; Riezman and Slemrod, 1987; Kenny and Winer, 2006; Aidt and Jensen, 2009). Moreover, there is a vast literature on the role of under-development in constraining tax structures both historically and in current developing countries.[9] Our theory proposes a micro-foundation that accounts for the changes in fiscal constraints over the course of development.

Recently, Besley and Persson (2008, 2009) propose an extension of the ratchet effect theory that emphasizes the role of increasing fiscal capacity over the course of development. They

[9]See, e.g., Alt (1983), Bird and Oldman (1990), Gillis (1989), Hettich and Winer (1991), Hinrichs (1966), Kelley and Oldman (1973), Kenny and Winer (2006), Webber and Wildavsky (1996).

develop a model where governments invest in fiscal capacity over time in response to wars. Historically, major wars have often been associated with government investments in tax capacity such as information reporting and tax withholding. While wars have undoubtedly been instrumental in the increased fiscal capacity of some countries such as the United Kingdom, other countries such as Sweden have experienced a smooth growth in its tax-to-GDP ratio that appears unrelated to wars (we discuss the empirical evidence in more detail in section 4). Furthermore, the question remains why recent (20th century) wars have lead to large government expansions, whereas earlier wars typically have not. Our paper contributes to this question and is therefore complementary to the Besley-Persson theory.

# 3 A Micro Theory of Third-Party Tax Enforcement

Let us assume that $N$ individuals are working in a firm and receive pre-tax wages $w = (w_1, .., w_N)$. The pre-tax profits of the firm are denoted by $\Pi$. Hence, the total value added created by the firm is equal to $V = W + \Pi$ where $W = \sum_n w_n$ are aggregate wages in the firm. Value added is also equal to total sales $S$ minus purchases $P$. Let us assume that the government imposes a flat tax at rate $\tau$ on both wages and profits. If $S$ and $P$ are observable to the government, then value added $V = W + \Pi = S - P$ is also observable. As a result, under-reporting wages is useless to the firm because this would automatically increase its tax on profits.[10] However, if $S$ and $P$ are not observable to the government, then the firm can possibly under-report wages $W$ without having to over-report profits $\Pi$.[11]

In practice, $S$, $P$, and $W$ (and hence $\Pi$) would be observable to the government if the firm truthfully records this information in its business records (such as accounting books and payroll lists) and the government has access to these business records. Some firms may be able to carry out their business without recording this information formally. For example, a small family business might carry out all or part of its purchases and sales with cash and never record this information. On the other hand, maintaining accurate business records is clearly helpful to firm productivity: the business can measure its profits accurately, keep track of wages paid out, plan production activities, obtain access to financial sector services, formal insurance, etc.

[10] If the tax rate on profits is lower than on wages, there is an incentive to under-report wages and over-report profits, and conversely.

[11] For example, the firm could exaggerate purchases or underreport sales. Symmetrically, the firm could under-report profits without having to over-report wages.

Realistically, the productivity gain of keeping business records is larger when the firm is larger and more complex, and for modern firms the cost of being off-the-books becomes prohibitive. We therefore assume that the firm maintains accurate business records, which creates potentially detectable information within the firm.[12] However, even though business records exist, the firm may still be able to hide those records from the government to evade taxes. For example, the firm may maintain a double set of books, true books for business purposes and edited books for tax purposes. In this section, we present a simple agency theory showing how the government can truthfully extract the true business record information using third-party reporting.[13]

Because we assume that the tax rate $\tau$ on profits and wages is the same, there are no incentives for profits and wage shifting and therefore wages and profits can be treated symmetrically. Hence, without loss of generality, we can model the owner of the profits as one additional wage earner, which simply amounts to ignoring profits (setting $\Pi \equiv 0$) in the analysis.[14]

## 3.1 Agency Model with Third-Party Reporting

### 3.1.1 Basic Setup

We assume that the government sets in place third-party reporting for tax purposes whereby each employee is required to report her earnings to the government and the firm is also required to report such individual earnings directly to the government.[15] Therefore, employees and employers have to agree on a wage report to the government as any discrepancy in the employer and employee reports would generate a tax audit.[16]

We can therefore assume that the firm and employees agree on reports to the government given by $\bar{w} = (\bar{w}_1, ..., \bar{w}_N)$, and this determines tax payments to the government unless any tax cheating is detected. We consider a situation where both real and reported wages $(w, \bar{w})$ are determined cooperatively by the $N$ employees of the firm. Because this is a tax collusion game, a

[12] In Section 4.4, we consider the implications of endogenizing the choice of being on the books as in Gordon and Li (2005).

[13] We focus primarily on third party reporting *within* the firm. We discuss briefly how third party reporting *between* firms, as happens with a value-added-tax, can also help enforcement.

[14] To be sure, in practice, profits are different from wages because they are not recorded in the same way. Wages are recorded on payroll lists while profits are typically obtained by substraction as $\Pi = S - P - W$.

[15] For example, in the United States, such reports are made through W2 forms issued by firms and sent to both the government and employees. Employees use this information to file their income tax returns (Logue and Slemrod, 2008 discuss this mechanism in detail). Some other OECD countries, such as Denmark, use pre-populated income tax returns whereby the government informs individuals about their earnings using information received from firms.

[16] Indeed, tax agencies systematically search for discrepancies between employee and employer reports to target tax audits.

cooperative game seems to be the most natural one.[17] As solution concept, we consider the core: no coalition of employees can break off from the firm and obtain strictly better outcomes for each member of this splitting coalition. In particular, the outcome of the cooperative game is Pareto efficient (otherwise the coalition of all employees could do better) and therefore maximizes total surplus of the employees in the firm. In this section, we take $N$ and the outside options of each employee as given. We denote by $\bar{y} = (\bar{y}_1, ..., \bar{y}_N)$ the disposable income levels (net of taxes) associated with those outside options.[18] In the general equilibrium macro-model presented in Section 4, we fully endogenize outside options and firm size $N$.

The presence of business records creates common knowledge within the firm. We capture such common knowledge by assuming that $(w, \bar{w})$ is known to everyone within the firm. In practice, although records may not be known to literally everyone within the firm, they are widely used in the firm and will be known by a number of employees. We explore also the alternative polar case where only employees for whom $w_n \neq \bar{w}_n$ are aware of tax evasion and can denounce tax cheating within the firm. This situation of private knowledge of tax evasion might be more realistic in the case of *external* parties such as business or individual clients, shareholders, or debt holders, a point we come back to later on.

Following the report $\bar{w}$ to the government, taxes are paid at rate $\tau$ based on $\bar{w}$. Each employee $n = 1, ..., N$ then decides either to stick to the report $\bar{w}_n$ or to whistleblow and reveal the true information to the government if $w \neq \bar{w}$. We further assume that internal business records create verifiable information: If any employee whistleblows and reveals the information $(w, \bar{w})$ of the company to the government and the government carries out an audit, the government will indeed be able to verify the information $(w, \bar{w})$ with the cooperation of the whistleblower. Because true business records are widely used within the company, it is impossible to hide them if a single knowledgeable insider is determined to reveal the true information to the government. In contrast, if no employee is willing the break a collusive tax cheating agreement, then it is much harder for the government to discover the true information. For simplicity, in that case, we assume that the government cannot detect cheating at all.

When evasion is detected, we assume that the government charges the evaded tax plus a

[17] The substance of our results generalizes to a non-cooperative game. The non-cooperative case always makes tax enforcement easier relative to the cooperative case.

[18] More precisely, we assume that outside options for any coalition of individuals is always given by $\bar{y} = (\bar{y}_1, ..., \bar{y}_N)$.

fine. As in all tax enforcement studies, we assume that there is an exogenous upper bound $\theta$ on the level of fines relative to tax evaded.[19] In that case, it is straightforward to show that it is always best for the government to impose the maximum possible fine in all circumstances. Therefore, without loss of generality, we assume that the penalty is equal to $\theta$ percent of the evaded tax to each person caught evading. In addition, the government may offer a reward to whistleblowers equal to a share $\delta$ of total uncovered tax evasion. For simplicity, we assume that all workers are risk neutral.[20]

The timing of the game is as follows: (1) employees agree cooperatively on a vector of wages $w = (w_1, ..., w_N)$ and a vector of reports $\bar{w} = (\bar{w}_1, ..., \bar{w}_N)$, (2) taxes are paid based on $\bar{w}$ at rate $\tau$, (3) each employee $n$ decides to stick to the report $\bar{w}_n$ or to whistleblow if $w \neq \bar{w}$, and (4) the government decides to audit or not, and fines and potential whistleblower rewards are paid.

**Proposition 1** *If all employees can commit ex-ante never to denounce tax cheating to the government, then in any cooperative equilibrium in the core, we have* $\bar{w}_n = 0$ *for all* $n$ *and no taxes are paid.*

**Proof:** Suppose that $\bar{w}_n > 0$ for some $n$. Then lowering $\bar{w}_n$ to zero increases the distributable surplus by $\tau \bar{w}_n$ and hence can increase the payoff of every employee without increasing the risk of detection as employees can commit not to denounce. Hence, $(w, \bar{w})$ with $\sum_n \bar{w}_n > 0$ cannot be in the core. QED.

The complete cheating equilibrium result of Proposition 1 is unlikely to be robust in practice. There are two sets of reasons why employees may denounce tax cheating to the government. The first set of reasons is the presence of *random shocks* such as a conflict between an employee and the employer, moral concerns of a newly hired employee, or simply a mistake whereby an employee reveals the true records $w$ to the government instead of the fake records $\bar{w}$. The second reason is the presence of rational whistleblowing if the government offers a reward to whistleblowers. We develop both models below and show that, when firms are large, the result of Proposition 1 is not robust as tax evasion is bound to be uncovered, which deters it in the

[19] Without such an upper bound, the government would impose infinite penalties and hence fully deter tax evasion in the first place. Such infinite fines are not tolerable in practice because punishment ought to be proportionate to the crime and because it is often very difficult to tell apart honest mistakes from intentional evasion. Therefore, imposing an upper bound on fines is both realistic and makes the tax enforcement theoretical problem non-trivial.

[20] Assuming risk aversion would make tax enforcement easier for the government. We consider risk aversion in Section 3.2 in the context of the Allingham-Sandmo model.

first place. As we shall see, the random shock model shows that the evasion equilibrium is not robust to introducing a *trembling hand*, while the whistleblower model shows that the evasion equilibrium is not robust to relaxing the *perfect commitment* assumption.

### 3.1.2 Random Shock Model

We incorporate the possibility that an employee may deviate and reveal internal business records either by mistake, because he is disgruntled, or because of moral concerns.[21] Let $\varepsilon$ be the probability of any given employee revealing true information through such random shocks. We assume for simplicity that those shocks are iid across employees. With $N$ employees, nobody will denounce tax cheating with probability $(1-\varepsilon)^N$. The probability that somebody in the firm reveals true information (and hence triggers an audit) is therefore given by $1-(1-\varepsilon)^N$. This probability is increasing in $N$, and tends to 1 as $N$ tends to infinity as a random shock is bound to happen when the number of employees is very large.

The expected pay-off of each employee equals

$$y_n = w_n - \tau \cdot \bar{w}_n - (1-(1-\varepsilon)^N) \cdot \tau \cdot (1+\theta) \cdot (w_n - \bar{w}_n)^+.$$

We assume that workers decide cooperatively on vectors of true and reported wages $(w, \bar{w})$, taking as given the random shocks in the second stage. The possible outcomes of this cooperative game (the core) are characterized by the set of vectors $(w, \bar{w})$ that maximize the total expected surplus $Y = \sum_n y_n$, subject to the resource constraint $\sum_{n=1}^N w_n = W$, non-negativity constraints $w_n, \bar{w}_n \geq 0$ for all $n$, and participation constraints $y_n \geq \bar{y}_n$ for all $n$, ensuring that each employee obtains a payoff that is at least as high as his best available outside option $\bar{y}_n$. The coalition of workers $1, ..., N$ will find it optimal to increase or decrease the report $\bar{w}_n$ for worker $n$ depending on the derivative of total surplus with respect to $\bar{w}_n$. When $\bar{w}_n < w_n$, we have:

$$\frac{\partial Y}{\partial \bar{w}_n} = \tau \cdot [-1 + (1+\theta)(1-(1-\varepsilon)^N)]. \tag{1}$$

When $\bar{w}_n > w_n$, we have: $\frac{\partial Y}{\partial \bar{w}_n} = -\tau$, so that it never pays to over-report wages.[22]

**Proposition 2** *In the random shock model, any cooperative solution is such that:*

[21] For example, an employee might no longer be able to condone tax cheating and decides to denounce the firm. Alternatively, a newly hired employee might not be willing to go along with tax cheating.

[22] In principle, in case of over-reporting uncovered by an audit, overpaid taxes will be refunded. This would not change the fact that $\partial Y / \partial \bar{w}_n < 0$ when $\bar{w}_n > w_n$.

(*a*) *If* $(1-\varepsilon)^N \leq \theta/(1+\theta)$*, there is no tax evasion at all:* $\bar{w} = w$.

(*b*) *If* $(1-\varepsilon)^N > \theta/(1+\theta)$*, there is complete tax evasion:* $\bar{w} = 0$.

(*c*) *For any* $\theta > 0$ *and* $\varepsilon > 0$*, there is* $\bar{N}$ *such as firms do not evade when* $N \geq \bar{N}$.

**Proof:** The proof of (a) and (b) is immediate as $\partial Y/\partial \bar{w}_n \geq 0$ iff $(1+\theta)(1-(1-\varepsilon)^N) \geq 1$ iff $\theta/(1+\theta) \geq (1-\varepsilon)^N$. For (b), where $\partial Y/\partial \bar{w}_n < 0$, the solution is determined by the non-negativity constraint $\bar{w}_n \geq 0$ for all $n$. For (c), $\bar{N}$ is defined by $\theta/(1+\theta) = (1-\varepsilon)^{\bar{N}}$, i.e., $\bar{N} = \log(\theta/(1+\theta))/\log(1-\varepsilon)$. QED.

Four points are worth noting about Proposition 2. First, when $\varepsilon = 0$, we are back to the standard collusive case where firm size does not help and there is always tax evasion. Second, when $\varepsilon > 0$ and even for moderate fines $\theta > 0$, it will always be the case that large firms choose not to evade, destroying the evasion equilibrium from Proposition 1. Our model can therefore explain why low fines and low audit rates can lead to successful enforcement in practice. This resolves the key puzzle of the Allingham-Sandmo model, which predicts extremely high evasion rates when audit rates and fines are low (given reasonable risk aversion parameters). Third, our qualitative results are robust to introducing risk aversion, which would make tax enforcement easier. Fourth, the results in the proposition do not depend on the specific division of revenue $W$ across workers. The equilibrium division will depend on the outside opportunities $\bar{y}$ and other factors not explicitly specified that determine the bargaining power of the individuals.

*Private vs. Common Knowledge of Cheating:*

The model above assumes that each employee has complete knowledge of the full set of wages $w, \bar{w}$. An alternative polar assumption is that each worker knows only about his/her own wages $w_n, \bar{w}_n$, while the employer is the only one knowing the full information $(w, \bar{w})$. This private knowledge model is more realistic in the case of *external* parties such as business or individual clients, shareholders, or debt holders, which share specific information with the firm but might not know the complete information within the firm. Critically, we maintain the assumption that, if there is under-reporting for individual $n$ ($\bar{w}_n < w_n$) and individual $n$ denounces the firm, the government will carry out an audit and then be able to observe the full set of actual and reported wages $w, \bar{w}$. This assumption can be defended as follows. A formal business needs to record $w$ and $\bar{w}$. Individual $n$ can prove that $w_n \neq \bar{w}_n$ as long as $w_n$ was formally paid out. Therefore, with hard evidence that the firm cheated on individual $n$, an investigation may

be able to retrieve the true business records and obtain full information $w, \bar{w}$. In other words, the firm is a nexus of information written in the internal business records, and the information cannot be broken or hidden into isolated pieces.

**Proposition 3** *In the random shock model with only private information on incomes:*
*(a) The optimal evasion strategy for the firm is to report zero income for the $N_c$ highest-paid employees, where $N_c$ is an integer below $\bar{N}$ defined as $[1-(1-\varepsilon)^{\bar{N}}](1+\theta)=1$.*
*(b) Assuming a fixed distribution of wage incomes, the fraction of income evaded tends to zero as $N$ gets large.*

**Proof:**
(a) If $N_c$ individuals evade, then the probability of detection equals $1-(1-\varepsilon)^{N_c}$ as only cheating individuals are able to denounce the firm. Hence, the total surplus is given by

$$Y=\sum_n [w_n-\tau\cdot\bar{w}_n-(1-(1-\varepsilon)^{N_c})\cdot\tau\cdot(1+\theta)\cdot(w_n-\bar{w}_n)^+].$$

When $\bar{w}_n < w_n$, we have:

$$\frac{\partial Y}{\partial \bar{w}_n}=\tau\cdot[-1+(1+\theta)(1-(1-\varepsilon)^{N_c})].$$

Therefore, evasion is profitable only if $N_c \leq \bar{N}$ defined as $[1-(1-\varepsilon)^{\bar{N}}](1+\theta)=1$. An equilibrium with $N_c \leq \bar{N}$ evaders Pareto dominates an equilibrium with truthful reporting, because the payoff from the $N_c$ evaders is higher due to underreporting, while the payoff from everybody else is unaffected. Moreover, when an employee evades, the surplus is maximized by full evasion: $\bar{w}_n = 0$. Because the extra surplus created by full evasion is proportional to $w_n$, surplus is maximized by having the highest-paid employees evade. Given $N_c \leq \bar{N}$, the optimal number of evaders reflects a trade-off between the extra surplus from the $N_c$th evader and the higher probability of being caught for all other evaders. It is optimal to evade for at least one employee (the highest paid) iff $\varepsilon\,(1+\theta)\leq 1 \Leftrightarrow \bar{N}\geq 1$.
(b) Because $\bar{N}$ is fixed, as $N$ goes to infinity, we have that $N_c/N \leq \bar{N}/N$ goes to zero—a vanishing fraction of employees will be able to evade. If the wage distribution is fixed, the share of total compensation going to a vanishing fraction of employees also converges to zero. QED.

Two points are worth noting about Proposition 3. First, our results of successful enforcement for large firms remains valid in the case of only private information, which is the least favorable

to tax enforcement. Second, this case may capture some of the real-world tax evasion practices of large firms. Most of the corporate income tax evasion does not take place as collusion to under-report the wages of ordinary employees, but takes place as under-reporting of profits by setting up illegal tax shelters. Such tax shelters are known or understood by a relatively small number of key accountants, a situation where the tax savings are large relative to the number of individuals in the know as in the proposition (see e.g., Slemrod 2004). Firms that plan on evading taxes therefore have an incentive to limit the flow of information within the firm.

### 3.1.3 Rational Whistleblower Model

We now consider the case where the government offers a whistleblower reward and we assume that each individual may voluntarily and rationally denounce their employer. Hence, we relax the critical assumption of ex-ante commitment from Proposition 1. In practice, firms do not have the power to enforce non-whistleblowing commitments.[23] We assume that the whistleblower reward is equal to a fraction $\delta$ of total uncovered revenue shared among all whistleblowers.[24]

Several OECD countries use such whistleblower rewards to induce insiders to denounce large-scale tax evasion within firms. For example, in the United States, the IRS Whistleblower Reward Program offers a payment of 15-30% of total uncovered tax revenue when whistleblowing leads to the detection of tax evasion in the excess of $2 million (Hesch, 2002). Related, Japan allows laid-off workers to claim unemployment benefits even if their employer did not pay social security contributions (OECD, 2004). Such claims help the government discover businesses evading social security taxes.[25] Alternatively, this model can be interpreted to capture moral rewards from denouncing large-scale tax cheating, assuming that each dollar of revenue that the whistleblower helps uncover creates a psychological reward of $\delta$ dollars.[26]

Given payments $w = (w_1, ..., w_N)$ and reports $\bar{w} = (\bar{w}_1, ..., \bar{w}_N)$, the payoff for employee $n$ if

[23]Organized crime can succeed in enforcing non-whistleblowing agreements by threats of severe retaliation. Short of falling into organized crime, firms cannot impose severe retaliation (Dixit, 2004). In a dynamic model, it is conceivable that whistleblowers could be fired and hence lose future rents from the employment match. Such an extension would make enforcement harder, but would not change the essence of our results.

[24]We discuss in Section 3.1.4 whether such a form of whistleblowing rewards can be seen as an optimal mechanism for the government to elicit tax compliance.

[25]Interestingly, laid-off employees no longer derive surplus from the employment relationship and hence have less to lose when denouncing tax evasion than current employees.

[26]If moral rewards are heterogeneous across individuals and unobservable by the employer, the model becomes conceptually very close to the random-shock model analyzed above.

he does not whistleblow is given by

$$y_n = w_n - \tau \bar{w}_n - a\,(1+\theta)\,\tau\,(w_n - \bar{w}_n)^+, \tag{2}$$

where $a = 0, 1$ is an audit dummy that takes the value 1 if any employee whistleblows. The payoff for employee $n$ if he whistleblows (in which case $a = 1$) is given by

$$y_n = w_n - \tau \bar{w}_n - (1+\theta)\,\tau\,(w_n - \bar{w}_n)^+ + \frac{\delta\,(1+\theta)\,\tau \sum_{n'} (w_{n'} - \bar{w}_{n'})^+}{N_w}, \tag{3}$$

where $N_w$ denotes the number of whistleblowers who share equally the rewards from whistleblowing. We assume that the whistleblower reward is a share of total revenue (including fines), because this turns out to be notationally simpler below.

From eqs (2)-(3), the total surplus in the firm can be written as

$$Y = \sum_{n'} \left[ w_{n'} - \tau \bar{w}_{n'} - a \cdot (1-\delta)\,(1+\theta)\,\tau\,(w_{n'} - \bar{w}_{n'})^+ \right]. \tag{4}$$

A cooperative solution $(w, \bar{w})$ maximizes surplus $Y$ subject to $\sum_{n'} w_{n'} = W$, non-negativity constraints $w_n, \bar{w}_n \geq 0$ for all $n$, and participation constraints $y_n \geq \bar{y}_n$ for all $n$. Notice that $(1-\delta)\,(1+\theta) \geq 1 \Leftrightarrow \ \delta \leq \theta/\,(1+\theta)$ is required to avoid a situation where employees always evade and then collectively whistleblow in order to recoup larger rewards than the fines they pay for under-reporting in the first place.

Moreover, because ex-ante commitments to not whistleblowing are infeasible, a cooperative solution with evasion must also satisfy incentive compatibility constraints ensuring that no worker finds it in his interest to whistleblow ex post. Therefore, given that co-workers do not whistleblow, utility for employee $n$ must be higher under no whistleblowing (eq. 2 with $a = 0$) than under whistleblowing (eq. 3 with $N_w = 1$), implying that, for all $n$,

$$\delta \leq \frac{(w_n - \bar{w}_n)^+}{\sum_{n'} (w_{n'} - \bar{w}_{n'})^+}. \tag{5}$$

On the other hand, if at least one co-worker whistleblows, employee $n$ will always find it in his interest to also whistleblow.

**Proposition 4** *In the whistleblower model, any cooperative solution is such that:*
*(a) If $N > 1/\delta$, then there can be no tax evasion at all: $\bar{w} = w$. Hence large firms do not evade taxes even if $\delta > 0$ is very small.*

*(b) If $N \leq 1/\delta$, then some evasion is sustainable, and an outcome without evasion is Pareto dominated by a sustainable evasion equilibrium. In the evasion equilibrium, the lowest-paid employee always reports zero wages (full evasion). All other employees may report positive wages (less than full evasion), but evade by at least as much as the lowest-paid employee in absolute terms. If wages $w_1, ..., w_N$ are equal, then all employees report zero wages.*

**Proof:** For (a), let us assume that $N > 1/\delta$ and that there is some evasion $E \equiv \sum_{n'}(w_{n'} - \bar{w}_{n'}) > 0$. Then, from eq. (5), we have $w_n - \bar{w}_n \geq \delta E$ for all $n$. Summing across all $n$, this implies $E \geq \delta \cdot N \cdot E$. Because $E > 0$, this implies $1 \geq \delta \cdot N$, which is a contradiction.

For (b), if some evasion is sustained ($E > 0$), then we must have $w_n - \bar{w}_n \geq \delta E$ for all $n$. Because $\delta \leq \frac{1}{N}$ in this case, it is feasible to satisfy this condition, for example by having equal evasion across all employees: $w_n - \bar{w}_n = \frac{E}{N} \geq \delta E$ for all $n$. Thus, starting from an outcome without evasion it is possible to reduce $\bar{w}_n$ by a small amount $d\bar{w}$ for all $n$ and thereby generate a sustainable Pareto improvement. The evasion equilibrium is characterized by the maximization of total surplus $Y$ at $a = 0$ subject to $\sum_{n'} w_{n'} = W$, non-negativity $w_n, \bar{w}_n \geq 0$, participation constraints $y_n = w_n - \tau \bar{w}_n \geq \bar{y}_n$, and the no-whistleblowing constraint (5) for all $n$. In this case, total surplus is given by $Y = (1 - \tau) W + \tau E$, implying that the equilibrium maximizes $E$ subject to $w_n - \bar{w}_n \geq \delta E$ and $w_n \geq 0, \bar{w}_n \geq 0$, $w_n - \tau \bar{w}_n \geq \bar{y}_n$ for all $n$. Because no employee can report negative wages, the no-whistleblowing constraint is hardest to satisfy for the lowest-paid individual, say employee 1, who can at the most evade by $w_1 = \min_n w_n \geq \bar{y}_n > 0$. Therefore, to maximize $E$, there is full evasion for the lowest-paid employee ($\bar{w}_1 = 0$) and total evasion is taken to the point where (5) is binding for this employee, $E = \frac{1}{\delta} w_1 \geq N w_1$. All other employees evade by at least as much as the lowest-paid employee in absolute terms, $w_n - \bar{w}_n \geq w_1$ for all $n$, but possibly by less in relative terms (less than full evasion). Obviously, if all wages are equal, then zero reporting by all employees is sustainable. QED.

Three points are worth noting about Proposition 4. First, if $\delta = 0$, i.e., if the government offers no reward for whistleblowing, then all firms will evade taxes as in Proposition 1. Second, as soon as some reward $\delta > 0$ is offered, then tax evasion is no longer sustainable for large firms. Therefore, the whistleblowing model also shows that low-powered fines and audit rates are enough to sustain truthful reporting in large firms. This shows that the collusion equilibrium of Proposition 1 is not robust to relaxing the assumption of perfect commitment. Third, in this model, equality in the distribution of true wages $w_1, ..., w_N$ has a positive impact on the

level of evasion that can be sustained in equilibrium. This is because low-paid workers are constrained in their evasion and therefore more tempted to whistleblow to get a share of total uncovered revenue. Because the wage structure is itself part of the cooperative evasion game, this creates an incentive for workers to agree on an equal wage structure so as to sustain full evasion. However, the equilibrium division of surplus depends also on the outside opportunities. In particular, complete wage equality and full tax evasion is not necessarily an equilibrium, because employees with good outside opportunities (presumably high-skilled workers) may not be willing to accept this division of surplus despite the extra tax evasion it delivers.

Finally, we may also consider the case with only private knowledge about cheating. Let us assume that only employees involved in cheating can denounce the firm, and that they form rational expectations about the extent of total cheating within the firm. Consistent with the random shock model, we would again have that the firm offers evasion to at most $N_c = 1/\delta$ employees, and cheating will be concentrated among the highest-paid employees. As $N$ becomes large, the fraction of employees evading and the share of total earnings evaded will shrink to zero.

### 3.1.4 Mechanism Design

The general lesson from our model is that common information among tax payers dramatically increases the ability of the government to extract tax revenue even with bounded fines. We have proposed a whistleblowing mechanism, which achieves perfect enforcement when $N$ is sufficiently large. The natural question is whether this mechanism is globally optimal, or if the government could do even better. Three points are worth noting.

First, when there is only one individual ($N = 1$) and keeping the assumption that the government can only successfully audit after whistleblowing, there is no mechanism that could induce the individual to reveal income truthfully.

Second, if there is more than one individual ($N \geq 2$), then in principle the government could design a non-conventional whistleblowing mechanism that induces truthful reporting. This mechanism is as follows: if the government receives information from $N_w$ whistleblowers, it will randomly select one whistleblower $n^*$, forgive $n^*$ his evaded tax and corresponding fine, and offer $n^*$ a small fraction of the tax evaded by the other individuals.[27] This mechanism would induce

[27]This mechanism is non-conventional in the sense that we are not aware of any tax agency implementing it in practice.

any individual to denounce tax cheating and make tax collusion impossible to sustain as long as $N \geq 2$. This strong implementation result is consistent with the mechanism design literature, which has shown that first best is often implementable in common information environments using sufficiently sophisticated mechanisms (Moore, 1992).

Third and most important, the complete enforcement result with a small number of individuals ($N \geq 2$) is not robust. An insider is willing to whistleblow only if rewards from whistleblowing are larger than the loss of breaking the collusion agreement. In our 1-period model and under the non-conventional mechanism described above, there is no loss from breaking collusion. However, in practice, breaking a tax collusion may generate both monetary costs (loss of future surplus from the worker-firm match, search costs to find a new job, etc.) and psychological costs (in the form of a conflict with colleagues). If those costs are non-trivial, then the net rewards from whistleblowing need to be non-trivial as well, and in this case evasion can only be fully deterred when $N$ is sufficiently large. Therefore, we believe that the results we have presented capture the gist of the real-world tax policy problem.

### 3.1.5 The Role of External Business Records and the Scope of the Firm

Our theory posits that the success of third-party reporting derives from the presence of verifiable *internal* business records that is commonly known among a sufficiently large number of employees. It is useful to contrast our theory with situations where such records are not present, or when *externally* recorded transactions allow outside business partners to denounce the firm.

*External Business Records and Value-Added Taxes*

Information on income generated by a business can also be obtained from external transactions. For example, businesses need to provide accounting records to shareholders or debt providers. Value added (equal to the sum of wages and profits as we discussed above) can be inferred from value added taxes (all OECD countries except the United States impose value added taxes).

The presence of publicly disclosed accounting books certainly imposes constraints on how much firms can evade as accounting books and corporate tax returns have to be consistent. Theoretically, the firm could collude with shareholders and banks to publicly disclose fake accounting books while secretly showing the true books to prospective shareholders and lenders. Exactly as in our model, such collusion would be very difficult to maintain with a large number of players. Therefore, firms which want to raise equity or debt need to maintain accurate

business records and cannot easily escape taxation.[28] If taxes on earnings are not linear, it is still possible to manipulate the distribution of reported earnings while truthfully reporting total earnings. This type of evasion could be analyzed along the lines we have proposed.

Value-added-taxes (VAT) require firms to keep accounts of all purchases and sales and pay taxes on sales net of purchases. Therefore, each firm has an incentive to under-report sales and over-report purchases hence creating opposite incentives across businesses engaged in arm-length transactions. Starting from a no evasion equilibrium, only businesses selling directly to households for final consumption can unilaterally evade by under-reporting sales. Even in that case, evasion is partial as businesses cannot consistently report negative value-added without raising suspicions. Exactly as in our model, we would expect small retailers to be able to evade partly the VAT while large retailing chains need to maintain formal business records making evasion much harder. Businesses further up in the VAT chain need to collude with businesses further down the chain to evade VAT. Therefore, as long as there is a large business further down the chain, VAT evasion is not feasible even for small informal businesses (Keen, 2007). However, if all businesses were small and informal, it would be impossible to implement a VAT as the tax would unravel from the bottom up. Therefore, in the end, we believe that it is again the presence of a large business which uses business records and cannot successfully hide them that makes the VAT successful, exactly as in our basic model.[29]

*Scope of the Firm*

Firms can evade some taxes by sub-contracting services, such as janitorial or building maintenance services, to providers which are often small and may not need to use business records. Such providers can evade taxes and therefore provide the service more cheaply than when those services are integrated and hence fully taxable. A particular example of such sub-contracting is given by tips, which are often additional off-the-books payments that take place directly between clients and employees. A related form of evasion takes the form of envelope wages where a share of wages is paid in cash outside the books. Such evasion is common in Eastern European countries in small businesses (OECD, 2004).

[28] As in Gordon and Li (2005), this debt channel is one of the benefits of using accounting books and being formal.

[29] No developing country with few large businesses can successfully implement a broad VAT (Ebrill at al., 2001). Furthermore, the VAT is not a necessary condition for successful corporate and individual income tax enforcement as show by the example of the United States.

## 3.2 Embedding Third-Party Reporting in the Allingham-Sandmo Model

The Allingham-Sandmo (A-S) model of tax evasion has been the dominating theoretical model for the analysis of tax evasion. It is therefore important to analyze how third-party reporting affects the results from this standard model.

Consider a taxpayer with true income $w = w_t + w_s$, where $w_t$ is subject to third-party reporting (such as wages and salaries) while $w_s$ is purely self-reported (such as self-employment income). The taxpayer reports $\bar{w}_t$ and $\bar{w}_s$ to the government. Based on our previous analysis, let us assume that we are in a situation with a large firm using business records, implying that third-party reporting is truthful and hence $\bar{w}_t = w_t$. Indeed, if $\bar{w}_t < w_t$, the tax administration will catch the under-reporting with probability one through matching of information returns and tax returns. Therefore, in this case, self-reported income $\bar{w}_s$ is the only choice variable of the individual as in the A-S model.

The probability of being audited and having one's under-reported income discovered is given by $p$. In the basic A-S model, the probability $p$ is independent of the reports $\bar{w}_s$ and $\bar{w}_t$. The taxpayer solves the following expected utility maximization problem:

$$\max_{\bar{w}_s} \quad (1-p) \cdot u\left(w - \tau(w_t + \bar{w}_s)\right) + p \cdot u\left(w\left(1-\tau\right) - \theta\tau\left(w_s - \bar{w}_s\right)^+\right). \tag{6}$$

### 3.2.1 Irrelevance of Third-Party Reporting

We can redefine the problem in terms of total reported income $\bar{w} \equiv \bar{w}_t + \bar{w}_s$, in which case the taxpayer maximizes with respect to $\bar{w}$:

$$\max_{\bar{w}} \quad (1-p) \cdot u\left(w - \tau\bar{w}\right) + p \cdot u\left(w\left(1-\tau\right) - \theta\tau\left(w - \bar{w}\right)^+\right). \tag{7}$$

This problem is exactly identical to the standard A-S problem with no third-party reporting. Let us denote by $w^*$ the level of earnings reported in the standard A-S model. We have:

**Proposition 5** *Irrelevance of third-party reporting in the standard Allingham-Sandmo model: Under a constant audit probability and with no constraints on self-reported income, the total level of earnings reported $\bar{w}_t + \bar{w}_s$ is equal to $w^*$. Therefore, tax evasion is independent of the fraction of income that is subject to third-party reporting.*

Third-party reporting does not improve enforcement because the taxpayer can counteract additional third-party reporting by adjusting self-reported income so as to achieve his optimal amount of evasion at the given probability of being detected and penalty. There is effectively 100% crowd-out of self-reported income by exogenously increased third-party reported income.

### 3.2.2 Breaking the Irrelevance Result

There are two assumptions underlying the strong result in Proposition 5: $(i)$ the tax policy allows losses to count fully against positive income in the tax calculation, $(ii)$ the enforcement policy imposes an audit probability that does not depend on the report $\bar{w}_s$, and in particular does not depend on whether $\bar{w}_s$ is positive or negative. In real-world tax systems, neither of these assumptions are typically satisfied. Deductibility of losses is disallowed to some degree and self-reported losses and deductions face higher audit rates than positive income. The above model offers a possible explanation of why this is so: it is because governments wish to protect the enforcement benefits of third-party reporting.

*Disallowing Losses*:

Let us assume that self-reported losses, $\bar{w}_s < 0$, are not allowed to be deducted for tax purposes from third-party reported income $\bar{w}_t > 0$. We prove the following proposition in appendix A.

**Proposition 6** *Suppose that losses are disallowed. Consider a small increase in third-party reported income $dw_t$ keeping total income $w$ constant.*
*$(a)$ If $w_t < w^*$, then $d\bar{w}_s/dw_t = -1$: increasing the fraction of income subject to third-party reporting does not increase total reported income $w_t + \bar{w}_s$ (100% crowd-out).*
*$(b)$ If $w_t > w^*$, then $d\bar{w}_s/dw_t = 0$: increasing the fraction of income subject to third-party reporting increases one for one total reported income $w_t + \bar{w}_s$ (0% crowd-out).*

In practice, governments often disallow losses in self-reported income to count against other income items. For example, the United States limits the deductibility of negative capital gains to only \$3,000.[30] Losses in passive activities are also not allowed to count against other positive income items. In those circumstances, the irrelevance result breaks down and third-party reporting does reduce overall tax evasion. We argue that those findings help explain the historical

[30]In the case of capital gains, although selling prices are often third-party reported, the buying price is self-reported in most cases, so that capital gains can be seen as effectively self-reported.

development of income taxation. Early income taxes started to raise significant revenue only after third-party income reporting became available and began as schedular taxes, i.e. with different income components being taxed separately. As a result, third-party reported tax bases were protected against losses in other self-reported tax bases. Initial comprehensive individual income taxes started with large exemption levels (so that only a minority of top income individuals were taxed) because, in early stages of economic development, many middle- and bottom-income earners were self-employed and income taxes cannot be enforced on such self-reported incomes. In contrast, third-party reporting was easier to implement for top incomes, which take the form of interest, dividends, large salaries, or business profits from large companies that are all recorded in formal accounting books.

*Differential Audit Rates*:

Let us assume that audit rates depend on self-reported income $\bar{w}_s$. We assume that $p(\bar{w}_s)$ is twice differentiable with derivatives $p'(\bar{w}_s) < 0$ and $p''(\bar{w}_s) \geq 0$. These assumptions appear to capture real-world auditing strategies in a stylized way. In particular, the assumption of a negative first-order derivative reflects that less aggressive/suspicious behavior (higher $\bar{w}_s$) lowers the risk of being audited.[31] The assumption of a non-negative second-order derivative reflects that, if the report is already unsuspicious (high $\bar{w}_s$) and the audit probability therefore low, increasing reported income will have a relatively small impact.[32]

**Proposition 7** *Suppose that the audit probability $p(\bar{w}_s)$ satisfies $p'(\bar{w}_s) < 0$, $p''(\bar{w}_s) \geq 0$. Consider a small increase in third-party reported income $dw_t$ keeping total income $w$ constant. Then $\frac{d\bar{w}}{dw_t} = 1 + \frac{d\bar{w}_s}{dw_t} > 0$: there is less than 100% crowd-out of increased third-party reported income by reduced self-reported income.*

The proof is presented in appendix A. Two points are worth noting. First, the condition $p''(\bar{w}_s) \geq 0$ is sufficient (but not necessary) to ensure incomplete crowding out.[33] Second, the key reason why there is less than 100% crowd-out is because the government uses third-party

[31] For example, in the United States, tax preparers are known to calibrate the audit probability to the wishes of their clients. An audit probability of about 0.3 is seen by tax preparers as an 'aggressive report'.

[32] For example, some countries select tax returns for audits based on computer-generated audit flags, where flags are triggered by return characteristics that appear to be suspicious or aggressive. A return may receive more than one flag if several line items raise suspicion, and the number of flags determine the probability of an audit. In such a system, as the return becomes more unsuspicious (reported income goes up) and the number of flags approaches zero, increases in reported income will have a very small impact on the audit probability.

[33] It is theoretically possible to obtain more than 100% crowding out if $p$ were (very) concave.

reports to determine its audit policy: $p$ is a function of $\bar{w}_s$ and hence of $w_t = \bar{w} - \bar{w}_s$ for a given $\bar{w}$. The case $p = p(\bar{w})$ would generate the same irrelevance result as in Proposition 5.

It would be useful to develop a formal normative tax and audit theory, which could generate as optimal policies the actual tax and audit policies observed over the course of economic development. We conjecture that disallowing self-reported losses could be desirable and that, more generally, the optimal audit policy would depend not only on total reported income but also on self-reported income. We leave this analysis for future research.

# 4 A Macro Theory of Tax Enforcement and Government Size

In this section, we set out a simple growth model that can explain the observed evolution of firm size, third-party income tax enforcement, and government size over the course of economic development. For expositional simplicity, we maintain the assumption that firms always maintain internal business records, which creates potentially detectable information within the firm. This assumption is not realistic for economies in very early stages of development where most firms are small and informal. We show in Section 4.4 that it is possible to endogenize the decision to use business records. In that case, endogenous books choice creates a consistent and reinforcing mechanism whereby growth and increasing firm size/complexity make it easier to enforce income taxation using third-party reporting.

## 4.1 Macro Model Without Enforcement Problems

*Households*

There is a continuum (of measure one) of homogeneous individuals, who derive utility $u(c, g)$ from the consumption of a private good $c$ and a tax financed public good $g$. We assume that $u(c, g)$ is homothetic, implying that the public good has an income elasticity equal to one (see below). We also assume that $u_c(c, 0) > 0$, so that public goods are not essential for prosperity.

We assume that labor is inelastically supplied. We denote by $w$ the pre-tax labor income of each individual and by $\tau$ the tax rate on income. Under truthful reporting, the budget constraint is given by $c = (1 - \tau)w$, where the price of the private good is normalized to one.

*Government*

We consider a benevolent government choosing public goods $g$ and taxes $\tau$ so as to maximize the welfare of the representative individual subject to a government budget constraint. The assumption of a benevolent government is not crucial for the model: as discussed earlier, our theory of government growth could alternatively be presented within the context of a Leviathan model where self-interested politician-bureaucrats maximize revenue for their own consumption.[34]

The government can convert one unit of $c$ into one unit of $g$. Absent any enforcement problem, the government budget constraint is given by $g = \tau w$. In this case, the government maximizes $u((1-\tau)w, \tau w)$ with respect to $\tau$, so that the standard Samuelson rule $u_c(c,g) = u_g(c,g)$ is satisfied. Because $u(c,g)$ is homothetic, the optimal tax rate $\tau^*$ is characterized by:

$$1 = \frac{u_g(c,g)}{u_c(c,g)} = \frac{u_g(1-\tau^*, \tau^*)}{u_c(1-\tau^*, \tau^*)}. \tag{8}$$

Importantly, the optimal tax rate is independent of income $w$ and hence will be constant along the growth path. Thus, optimal government spending as a share of income, $g^*/w = \tau^*$, is constant and the public good income elasticity is equal to one. This implies that the size of government to GDP would be constant over time in the absence of enforcement problems.

*Firms and Productivity*

We assume that all firms have access to the same production technology. For each firm, the average product of labor equals $x(N,A)$, where $N$ is the number of employees in the firm and $A$ is a technology parameter that grows exogenously over time. We assume that $x(N,A)$ is increasing in $A$ and inversely U-shaped in $N$. The assumption that average productivity is inversely U-shaped in $N$ mirrors the standard assumption of a U-shaped average cost curve. Furthermore, we assume that technological progress is complementary to labor input, defined as $x_A(N,A)/x(N,A)$ being increasing in $N$.

Let $\hat{N}(A)$ be the firm size maximizing average productivity (minimizing average costs), i.e. $\hat{N}(A) \equiv \arg\max_N \; x(N,A)$. This implies $x_N(\hat{N},A) = 0$ and $x_{NN}(\hat{N},A) < 0$. We then have

$$\frac{d\hat{N}}{dA} = -\frac{x_{NA}(\hat{N},A)}{x_{NN}(\hat{N},A)} > 0,$$

where the inequality follows from the assumption $x_A/x$ increasing in $N$ (and using $x_N(\hat{N},A) = 0$), which implies $x_{AN}(\hat{N},A) = x_{NA}(\hat{N},A) > 0$.

[34] Although both models can provide a positive theory of government growth, their normative implications are obviously very different.

We assume perfect competition in all markets, implying that firms take the output price and wages as given. Profits are given by $x(N, A) \cdot N - w \cdot N$, which is maximized with respect to firm size $N$. The first-order condition for firm size is given by $x_N \cdot N + x - w = 0$. We assume that there is free entry of firms, which leads to zero profits in general equilibrium. Hence, we have $x = w$ and the first-order condition for $N$ reduces to $x_N(N, A) = 0$. Therefore, the optimal size of firms is given by the productivity-maximizing level $\hat{N}(A)$.

In our model, $N$ is the number of employees in the firm so that we can directly apply the model from Section 3. It would also be possible to interpret $N$ more broadly as the number of external parties that share some of the information of the business. In such an interpretation, a more inter-connected production process becomes more valuable as technology progresses.

## 4.2 Incorporating Tax Evasion into the Model

We consider the whistleblower model of tax evasion. The whistleblower model simplifies the presentation, because it involves no uncertainty. From Proposition 4, either there is evasion that always goes undetected or there is no evasion at all. Furthermore, because all workers are identical in this model, when there is evasion, it is complete.

As before, we consider a cooperative game where the firm and its employees agree on true and reported wages $(w, \bar{w})$ to maximize total surplus. Either they report truthfully $(\bar{w} = w)$ and workers pay taxes $\tau w$, or they report dishonestly $(\bar{w} = 0)$ and workers pay no tax. For expositional simplicity, it is convenient to assume that the firm has all the bargaining power, implying that the solution maximizes profits under the constraint that each employee receives his outside option. Therefore, unlike the micro model in Section 3, we do not characterize the entire set of cooperative equilibria (the core), but a specific equilibrium where the firm gets the surplus from evasion.[35] Notice though, that in general equilibrium where free entry eliminates pure profits, the workers ultimately receive all the surplus from tax evasion.

Let $\bar{y}$ be the net-of-tax income of each employee in his best outside option, where $\bar{y}$ is determined by the equilibrium in the labor market and taken as given by the firm. The firm has to offer each employee a pre-tax compensation equal to $\bar{y}/(1 - \tau)$ if it complies with the tax law, and equal to $\bar{y}$ if it evades all taxes. Denoting by $1(\bar{w} = w)$ the indicator variable

[35] This equilibrium is natural given the assumptions of no hiring-firing costs and perfect competition in the labor market. Under those assumptions, if one worker does not accept the proposed division of surplus, the firm can costlessly hire another worker at his marginal product.

equal to one under truthful reporting and zero under full evasion, profits can be written as $x(N,A) \cdot N - \frac{\bar{y}}{1-\tau \cdot 1(\bar{w}=w)} \cdot N$. Hence, for the firm, under-reporting wages to the government lowers the before-tax wage it has to pay its employees. The potential cost of under-reporting is that it may be denounced by an employee seeking the whistleblower reward $\delta$.

If the firm does not evade, then we saw in the previous section that equilibrium firm size equals $\hat{N}(A)$, the before-tax wage is given by $w = x(\hat{N}(A), A)$, and the after-tax wage is given by $y = (1-\tau) \cdot x(\hat{N}(A), A)$. If the firm evades and nobody whistleblows, each employee income is $w = y = x(N, A)$. If an employee whistleblows (and nobody else does), he obtains $x(N,A) - \tau(1+\theta)x(N,A) + \delta\tau(1+\theta)x(N,A)N$. Therefore, the employee does not whistleblow iff $x(N,A) \geq x(N,A) - \tau(1+\theta)x(N,A) + \delta\tau(1+\theta)x(N,A)N$, which is equivalent to $N \leq 1/\delta$ as in Proposition 4. Hence, a firm that evades tax has to choose a firm size below $1/\delta$.[36]

**Proposition 8** *We obtain the following cases:*

*(1) If $\hat{N}(A) \leq 1/\delta$, then the firm evades all taxes and chooses the optimal firm size $\hat{N}(A)$.*

*(2) If $\hat{N}(A) > 1/\delta$ then:*

*(a) If $x(\hat{N}(A), A) \cdot (1-\tau) < x(1/\delta, A)$, then the firm evades all taxes and chooses sub-optimal firm size $1/\delta$.*

*(b) If $x(\hat{N}(A), A) \cdot (1-\tau) \geq x(1/\delta, A)$, then the firm does not evade and chooses the optimal firm size $\hat{N}(A)$.*

**Proof:** The proof of (1) follows from the fact that profits are always greater under evasion when this can be sustained at the optimal firm size $\hat{N}(A)$. The proof of (2a) and (2b) follows from the observation that, once evasion is not sustainable under the optimal firm size $\hat{N}(A)$, an evading firm must reduce firm size to $1/\delta$. Under full evasion and $N = 1/\delta$, the free-entry (zero-profit) equilibrium is characterized by labor income $y = x(1/\delta, A)$. Under no evasion and $N = \hat{N}(A)$, the free-entry equilibrium has labor income $y = (1-\tau)x(\hat{N}(A), A)$. In a labor market equilibrium, the outcome will be the one associated with the highest labor income, which gives the conditions in the proposition. QED.

[36] Notice that the decision to whistleblow is independent of the level of public goods $g$, because whistleblowing within a single firm does not affect the aggregate level of $g$.

Note that Proposition 8 implies that taxation distorts firm size away from intermediate levels above $1/\delta$. The result is consistent with the empirical phenomenon of the "missing middle" discussed in the development literature (e.g., Tybout, 2000). A recent paper by Dharmapala, Slemrod and Wilson (2008) argues that the missing middle may be the outcome of optimal tax policies that exempt small firms from taxation in order to save on administrative costs. In our model, the missing middle does not arise because small firms are tax exempt *de jure*, but because small firms can sustain tax evasion and therefore become tax exempt *de facto*.

## 4.3 Macroeconomic Development and Optimal Government Policy

We now turn to the evolution of government size over the growth process. Let $A_L$ be the technology level such that $\hat{N}(A_L) = 1/\delta$ and $A_H$ the technology level such that $x(\hat{N}(A_H), A_H) \cdot (1 - \tau^*) = x(1/\delta, A_H)$. Obviously, we have $0 < A_L \leq A_H$ and $A_L = A_H$ iff $\tau^* = 0$.

**Proposition 9** *We have the following three stages of development:*
*(1) Early Stage: when $A \leq A_L$, the government cannot raise any tax revenue and sets $\tau(A) = 0$.*
*(2) Intermediate Stage: when $A_L < A < A_H$, the government is constrained by tax enforcement and sets $\tau(A)$ such that $x(\hat{N}(A), A) \cdot (1 - \tau(A)) = x(1/\delta, A)$. Firms do not evade taxes. Government size is suboptimal, $\tau(A) < \tau^*$, and $\tau(A)$ is increasing in $A$.*
*(3) Late Stage: when $A \geq A_H$, the government is no longer constrained by tax enforcement and firms do not evade taxes. The tax rate is set at the optimal level $\tau(A) = \tau^*$ and government size is constant in $A$.*

**Proof:** The only non-obvious point is that $\tau(A)$ increases in $A$ in the intermediate stage. Log-differentiating $1 - \tau(A) = x(1/\delta, A)/x(\hat{N}(A), A)$ and using $x_N(\hat{N}, A) = 0$, we obtain

$$-\frac{1}{1-\tau(A)}\frac{d\tau(A)}{dA} = \frac{x_A\left(1/\delta, A\right)}{x\left(1/\delta, A\right)} - \frac{x_A(\hat{N}(A), A)}{x(\hat{N}(A), A)},$$

Because $\hat{N}(A) > 1/\delta$ in the intermediate stage, the assumption that technological progress is complementary to labor input, $x_A/x$ increasing in $N$, implies $d\tau/dA > 0$. QED.

The predictions of Proposition 9 are illustrated in Figure 1. Following an early stage with zero tax revenue and no public goods provision, the government gradually increases the tax rate over the growth process until it reaches the dashed line in the figure after which government size as a share of income is constant. This simple model captures the stylized facts regarding the

evolution of taxes and government size over the process of economic development. Indeed, all of the advanced economies in the world experienced a drastic increase in the size of government to GDP during the 20th century when broad-based and third-party enforced taxes such as individual income taxes, payroll taxes, and value-added-taxes were gradually implemented (see, e.g., Webber and Wildavsky, 1986). In the last three-four decades, the size of government to GDP has been roughly constant in the richest economies (OECD, 2008).

It is illuminating to compare our theoretical predictions to the empirical evidence by considering the historical evolution of taxation and government size in three different advanced economies. Figure 2 (Panel A) displays the ratio of tax revenue (including all social security contributions and all levels of government) to GDP in the United States, the United Kingdom, and Sweden since the latter part of the 19th century. In all three countries, the tax ratio is low (well below 10%) and very flat until World War I, increases until around the late 1970s, and then stays roughly constant thereafter. The exact timing of the tax increases and the final level of the ratio differ across countries. Most of the increase take place around the World Wars in the United Kingdom. The United States also displays clear spikes around the World Wars, although the tax ratio comes down to some extent after the wars. The increase in government size is a lot smoother in Sweden, which was relatively unaffected by the wars due to its status as a neutral country. However, in all three countries and despite their different exposure to wars, the stylized pattern of government growth is the same and fits very nicely with the theoretical prediction shown in Figure 1. The case of Sweden is important to show that external shocks and the ensuing ratchet effects are not necessary for the growth of government. Indeed, an examination of all 30 OECD countries (OECD, 2008) shows that only Mexico and Turkey—the two poorest OECD countries—have tax-to-GDP ratios below 25% in 2006, implying that the growth of the tax-to-GDP ratio is universal among advanced economies.

Panel B decomposes the US tax ratio into incomes taxes (individual, corporate, and all payroll taxes) and other taxes (property taxes, excise and sales taxes, custom duties, etc.). Income taxes correspond roughly to modern taxes (based on business records, third-party reporting, etc.), while other taxes are traditional taxes. Consistent with our theory, the graph shows clearly that the secular growth in the tax ratio comes from income taxes, with only very modest increases in other taxes. The US case is representative of other OECD countries where income

and payroll taxes, along with the value-added tax,[37] account for the bulk of government growth. A more comprehensive analysis of tax ratios and composition over time and across countries is left for future research. Finally, note that it is easy to extend our model to incorporate archaic taxes by assuming that the government can raise a fixed fraction $\tau_0$ of national product through such taxes. In that case, the theoretical path of taxes to GDP depicted on Figure 1 would be shifted upward by $\tau_0$ exactly as on Panel B in Figure 2.

## 4.4 Extensions

### 4.4.1 Endogenous Use of Business Records

Our analysis has assumed the existence of accurate business records ('books') that create potentially detectable evidence of tax evasion. One way for a firm to escape taxation completely is to discard the use of books altogether. As discussed in Section 3, being off-the-books is presumably associated with a productivity loss that is growing in firm size and complexity, and firms choose to be on or off books by trading off this productivity loss against the tax savings as in Gordon and Li (2005). It is conceptually straightforward to set out a macro-economic model along these lines, which generates results that are fully consistent with those presented above.

We assume that the average product of labor for a firm on the books is equal to $x(N, A)(1-c)$, where $c$ is a fixed administrative cost of maintaining books per unit of output and $x$ has the same properties as in the earlier model. Average productivity for a firm off-the-books equals $x(N, A)\alpha(N)$, where $\alpha$ reflects the output loss of not having accurate business records. We make the assumptions $\alpha(0) = 1$, $\alpha'(N) < 0$, and $\lim_{N\to\infty} \alpha(N) = 0$.[38] To zoom in on the implications of endogenous books, we drop the agency model of third-party reporting, and assume simply that a firm on the books is perfectly taxable while a firm off the books is completely untaxable. All other components of the model (such as the specification of consumers and government) are exactly as before. In this setup and under some additional regularity assumptions, it is possible to state a result analogous to Proposition 9 and with an evolution of government size as in Figure 1. The mechanism that drives this development is no longer the increased risk of third-party whistleblowing but rather an increased productivity gain of using rigorous

[37] The value-added-tax relies critically on accounting books and third party enforcement is obtained through arm-length business-to-business transactions.

[38] In other words, not using books becomes prohibitively costly as technological progress grows. The results easily extend to the case where $\lim_{N\to\infty} \alpha(N) = \bar{\alpha} > 0$ as long as $\tau^* \leq 1 - \bar{\alpha}/(1-c)$, i.e., the social optimal tax rate is not too large.

business records as firms get larger. The model and results are presented in appendix B.1.

### 4.4.2 Endogenous Growth

The above analysis of the development of tax enforcement and government size assumes that productivity increases exogenously. This is a reasonable assumption if government activities have only a limited impact on the growth process. However, some government activities, such as the protection of property rights, law enforcement, and investments in education, health, and infrastructure, may be very important for growth. Barro (1990) develops an endogenous growth model where government inputs are complementary to private inputs in production, and derives the optimal tax rate and government size along the growth process. It is possible to embed a Barro-type production technology in our theory of tax enforcement, and obtain a model where optimal firm size grows with endogenous technological progress. We present this model in appendix B.2. The government applies a time varying tax rate to maximize the lifetime infinite-horizon utility of a representative household. Under some parameters, an economy might get stuck in a poverty trap, because firms are too small and the government cannot raise income taxes to feed the growth process. When the economy is not stuck in a poverty trap, there will be three stages of development as in Proposition 9. First, the government cannot raise income taxes and the economy grows too slowly relative to first best. Second, the government starts raising income taxes, but the tax rate is constrained by tax enforcement. The economy grows faster but still slower than first best. Third, the government is no longer constrained by tax enforcement and can apply the tax rate that optimizes the growth rate. Thus, this endogenous growth model delivers the same S-shaped time pattern of the tax-to-GDP ratio that fits the empirical evidence. The model also suggests that the inability of some of the poorest countries to start the growth process might be due to insufficient fiscal capacity.[39]

# 5 Conclusion

Our paper has presented a simple agency model to explain why third-party income reporting by employers can sustain tax enforcement in spite of low fines and low audit rates. Therefore, our model overcomes the main shortcoming of the standard Allingham-Sandmo model of tax evasion.

[39]Economists have proposed many theoretical mechanisms that may generate poverty traps (see Azariadis and Stachurski, 2005, for a survey). The public finance theory described above should be seen as complementary to those alternative theories.

The key mechanism that makes third-party tax enforcement successful is the combination of verifiable book evidence that is common knowledge within the firm and a large number of employees, as any single employee can denounce collusive tax cheating between employees and the employer by revealing true books to the government.[40] Embedding this agency model into the standard Allingham-Sandmo tax evasion model, we have shown that third-party reporting improves tax enforcement if the government disallows self-reported losses or audits such losses more stringently. We have argued that those findings fit very well with actual tax policy practices and help explain the historical development of income taxation. We have also embedded this agency model into a macro-economic growth model where the size and complexity of firms grows with technological progress. Our simple model can capture the stylized fact that the level of income taxes and the size of government to GDP grows during the process of economic development. In our model, capitalistic development relaxes the tax enforcement constraint and naturally leads to large welfare state governments.

Our analysis has been primarily theoretical. In future work, it would be valuable to test some of the predictions of our model both in developed and developing countries.

Our theory predicts that third-party enforcement should be most successful for large and complex firms. The related theories proposed by Gordon and Li (2005) and Kopczuk and Slemrod (2006) point out that links to the financial sector and the network of arm-length transactions between firms (respectively) explain the success of modern taxes. We think that both internal common knowledge (as in our model) and external arm-length transactions (as in Kopczuk and Slemrod) produce verifiable information that the government can exploit for tax purposes. Hence, it is really the volume of recorded transactions (both internal and external) that grows with economic development and increases the ability to tax. In principle, an empirical analysis of tax audits of both firms and employees in a developed country could be used to assess which factors—size and complexity, links to the financial sector, network of transactions—explains best the low levels of tax evasion observed in advanced OECD countries.

Our theory also predicts that the inability to collect income taxes from the informal sector is the key reason why developing countries collect little tax revenue.[41] Other theories have

[40]It is an intriguing question whether the development of automatic tax withholding and tax return free systems could affect this mechanism as employees may no longer have to certify or even be aware of what employers report to the government.

[41]The theory proposed by Gordon and Li (2005) makes the same prediction.

been put forward: (1) corruption in the tax administration may make taxes hard to collect in both the formal and informal sectors, (2) demand for government services may be lower in poor countries. We could test our theory by estimating tax rates in the formal and informal sectors of developing countries and comparing them with tax rates in OECD countries. Our theory predicts that tax rates on the formal sector in developing countries should be high—possibly as high as in OECD countries—while the alternative theories imply that even in the formal sector, tax rates should be much lower in developing countries than in OECD countries.

# A Appendix: Proofs for the A-S Model

*Proof of Proposition 6*

Suppose self-reported losses are disallowed to count against third-party reported income for tax purposes. In this case, the taxpayer maximizes

$$(1-p)\cdot u\left(w-\tau(w_t+\bar{w}_s^+)\right)+p\cdot u\left(w-\tau(w_t+w_s^+)-\theta\tau\left(w_s^+-\bar{w}_s^+\right)\right), \tag{9}$$

with respect to $\bar{w}_s$. Now we have to distinguish explicitly between positive self-reporting ($\bar{w}_s > 0$) and non-positive self-reporting ($\bar{w}_s \leq 0$). As losses are disallowed, individuals with non-positive self-reporting ($\bar{w}_s \leq 0$) are indifferent on how much losses to report, so we assume without loss of generality that $\bar{w}_s = 0$ in that case. Therefore $\bar{w}_s^+ = \bar{w}_s$. Positive self-reporting is optimal if the right-derivative of (9) is positive at $\bar{w}_s = 0$, i.e. if

$$\frac{u'\left(w-\tau w_t-\theta\tau w_s^+\right)}{u'\left(w-\tau w_t\right)} > \frac{1-p}{p\theta}, \tag{10}$$

which just corresponds to saying that the solution to the original problem was associated with $w^* > w_t$ (i.e., $\bar{w}_s > 0$). If equation (10) holds, the maximization problems with symmetric versus asymmetric treatment of gains and losses yield the same solutions, and third-party reporting is irrelevant. In other words, third-party reporting on infra-marginal units of reported income (increasing $w_t$ when $w_t < w^*$) has no effect on $\bar{w}$. On other hand, once third-party reporting starts digging into previously unreported income (increasing $w_t$ when $w_t \geq w^*$), then $\bar{w}_s = 0$ and the individual can no longer offset third party reporting. QED.

*Proof of Proposition 7*

The taxpayer maximizes

$$\left(1-p\left(\bar{w}_s\right)\right)\cdot u\left(w-\tau(w_t+\bar{w}_s)\right)+p\left(\bar{w}_s\right)\cdot u\left(w\left(1-\tau\right)-\theta\tau\left(w_s-\bar{w}_s\right)\right). \tag{11}$$

Denoting by $c_A$ and $c_N$ disposable income in the audit and no-audit states, the first-order condition for $\bar{w}_s$ is given by

$$p\left(\bar{w}_s\right) u'\left(c_A\right) \theta\tau - p'\left(\bar{w}_s\right) \left[u\left(c_N\right) - u\left(c_A\right)\right] = \left(1 - p\left(\bar{w}_s\right)\right) u'\left(c_N\right) \tau, \tag{12}$$

where the left-hand side is the gain of a higher report (the higher utility in the audited state plus the lower probability of being in the audited state), while the right-hand side is the cost of a higher report (the lower utility in the non-audited state).

Let us rewrite the first-order condition (12) as

$$0 = R(\bar{w}_s, \bar{w}, w) \equiv p(\bar{w}_s)\theta\tau u'(c_A) - (1 - p(\bar{w}_s))\tau u'(c_N) - p'(\bar{w}_s)[u(c_N) - u(c_A)], \tag{13}$$

where $c_N = w - \tau\bar{w}$ and $c_A = w(1-\tau) - \theta\tau(w - \bar{w})$ are functions of $w$ and $\bar{w}$ (and not of $\bar{w}_s$). Since $\bar{w} = \bar{w}_s + w_t$ and the individual takes $w_t$ as fixed, the second-order condition of the individual maximization problem is given by

$$\frac{dR}{d\bar{w}_s} = \frac{\partial R}{\partial \bar{w}_s} + \frac{\partial R}{\partial \bar{w}} < 0.$$

Consider now a change in third-party reporting $dw_t$ keeping $w$ constant such that $dw = 0$ and $d\bar{w}_s = d\bar{w} - dw_t$. Differentiating (13), we obtain

$$0 = \frac{\partial R}{\partial \bar{w}_s}[d\bar{w} - dw_t] + \frac{\partial R}{\partial \bar{w}} d\bar{w},$$

which implies

$$\frac{d\bar{w}}{dw_t} = \frac{\partial R/\partial \bar{w}_s}{\partial R/\partial \bar{w} + \partial R/\partial \bar{w}_s}.$$

The denominator is negative from the second-order condition. Hence $d\bar{w}/dw_t > 0$ if and only if $\partial R/\partial \bar{w}_s < 0$. As $c_A, c_N$ do not depend on $\bar{w}_s$, we have indeed

$$\frac{\partial R}{\partial \bar{w}_s} = p'(\bar{w}_s)\tau[\theta u'(c_A) + u'(c_N)] - p''(\bar{w}_s)[u(c_N) - u(c_A)] < 0,$$

where the inequality follows from $p'(\bar{w}_s) < 0$, $p''(\bar{w}_s) \geq 0$, and $c_N \geq c_A$. QED.

# B Appendix: Extensions

## B.1 Endogenous Use of Business Records

We assume that all firms have access to the same production technology. For each firm, the average product of labor equals $F(N, A, B)$, where $N$ is the number of employees in the firm, $A$ is a technology parameter that grows exogenously over time, and $B$ is an indicator variable that equals 1 when the firm uses books and equals 0 otherwise. As in Section 4.4.1, we assume

$$F(N, A, B) = \begin{cases} x(N, A)(1 - c) & \text{for } B = 1 \\ x(N, A) \cdot \alpha(N) & \text{for } B = 0 \end{cases},$$

where $x(N, A)$ is increasing in $A$ and inversely U-shaped in $N$ (as in Section 4), $c$ is a resource cost in proportion of output of bookkeeping, while $1 - \alpha(N)$ reflects the output loss of not using books. We assume that $\alpha(0) = 1$, $\alpha_N(N) < 0$, and $\lim_{N\to\infty} \alpha(N) = 0$.[42] Let $\hat{N}(A) = \arg\max x(N, A)$. As in Section 4, we assume that technological progress is complementary to labor input, defined as $x_A(N, A)/x(N, A)$ being increasing in $N$. This implies that $\hat{N}'(A) > 0$. Moreover, we assume that, for all $N$,

$$\lim_{A\to 0} \hat{N}(A) = 0, \ \lim_{A\to\infty} \hat{N}(A) = \infty, \ \lim_{A\to\infty} \frac{x(N, A)}{x(\hat{N}(A), A)} = 0. \tag{14}$$

Under those assumptions, we can prove:

**Proposition 10** *There are two cut-off levels $A_L < A_H$ which determine three stages of development:*

*(1) Early Stage: when $A \leq A_L$, firms do not use books and the government cannot raise any tax revenue and sets $\tau(A) = 0$.*

*(2) Intermediate Stage: when $A_L < A < A_H$, the government is constrained by tax enforcement. Firms use books and do not evade taxes. Government size is suboptimal, $\tau(A) < \tau^*$, and $\tau(A)$ is increasing in $A$.*

*(3) Late Stage: when $A \geq A_H$, the government is no longer constrained by tax enforcement and firms do not evade taxes. The tax rate is set at the optimal level $\tau(A) = \tau^*$ and government size (relative to total product) is constant in $A$.*

[42] Note that the assumption $\lim_{N\to\infty} \alpha(N) = 0$ does not necessarily imply that output vanishes for large $N$ since output equals $x(N, A) \cdot \alpha(N) \cdot N$.

**Proof:** Firm profits $\pi(N, A, B)$ are such that

$$\pi(N, A, 0) = x(N, A)\,\alpha(N)\,N - yN, \tag{15}$$

$$\pi(N, A, 1) = x(N, A)\,(1-c)\,N - \frac{y}{1-\tau}N, \tag{16}$$

where $y$ is the net-income that the firm has to offer its employees, while $\tau$ is the tax rate on earnings when using books. Profits are maximized with respect to $N$ and $B$. The first-order conditions with respect to $N$ equals

$$x_N(N, A)\,N\alpha(N) + x(N, A)\,\alpha_N(N)\,N + x(N, A)\,\alpha(N) - y = 0 \text{ for } B = 0, \tag{17}$$

$$[x_N(N, A)\,N + x(N, A)]\,(1-\tau)\,(1-c) - y = 0 \text{ for } B = 1. \tag{18}$$

Let us denote by $N(A, 0)$ and $N(A, 1)$ the optimal choices for $N$ given by (17) and (18). There is free entry/exit of firms and labor is completely mobile across firms. This implies that profits are zero in equilibrium. From the profit expressions (15) and (15), we get

$$y = x(N(A, 0), A) \cdot \alpha(N(A, 0)) \qquad \text{for } B = 0, \tag{19}$$

$$y = x(N(A, 1), A)\,(1-\tau)\,(1-c) \qquad \text{for } B = 1. \tag{20}$$

Using these two expressions, the first-order conditions (17) and (18) simplify to

$$x_N(N(A, 0), A)\,\alpha(N(A, 0)) + x(N(A, 0), A)\,\alpha_N(N(A, 0)) = 0 \text{ for } B = 0, \tag{21}$$

$$x_N(N(A, 1), A) = 0 \text{ for } B = 1. \tag{22}$$

Comparing these two expressions, we see that a firm with books will choose more employees than a firm without books:

$$N(A, 1) > N(A, 0). \tag{23}$$

**Lemma 1** *Our assumption of complementarity implies*

$$\frac{dN(A, B)}{dA} > 0 \text{ for } B = 0, 1.$$

**Proof of the Lemma:** For $B = 1$, we have from (22)

$$\frac{dN(1)}{dA} = -\frac{x_{NA}(N(1), A)}{x_{NN}(N(1), A)} > 0,$$

which is positive because of the assumption of complementarity and because $x(N, A)$ is inversely U-shaped in $N$.

For $B = 0$, the first-order condition (21) is

$$\Phi\left(A, N\left(0\right)\right) = x_N\left(N\left(0\right), A\right)\alpha\left(N\left(0\right)\right) + x\left(N\left(0\right), A\right)\alpha_N\left(N\left(0\right)\right) = 0. \tag{24}$$

At the optimum, we have

$$\frac{dN\left(0\right)}{dA} = -\frac{\Phi_A\left(A, N\left(0\right)\right)}{\Phi_N\left(A, N\left(0\right)\right)},$$

where $\Phi_N\left(A, N\left(0\right)\right) < 0$ because of the second-order condition. This implies

$$\text{sign}\left[dN\left(0\right)/dA\right] = \text{sign}\left[\Phi_A\left(A, N\left(0\right)\right)\right].$$

From (24), we have

$$\Phi_A\left(A, N\left(0\right)\right) = x_{NA}\left(N\left(0\right), A\right)\alpha\left(N\left(0\right)\right) + x_A\left(N\left(0\right), A\right)\alpha_N\left(N\left(0\right)\right),$$

where we have used the Envelope Theorem. By inserting (24), we see that

$$\text{sign}\left[\Phi_A\left(A, N\left(0\right)\right)\right] = \text{sign}\left[x_{NA}\left(N\left(0\right), A\right) - \frac{x_A\left(N\left(0\right), A\right)x_N\left(N\left(0\right), A\right)}{x\left(N\left(0\right), A\right)}\right],$$

which is positive because of the complementarity assumption $x_A\left(N, A\right)/x\left(N, A\right)$ increasing in $N$. QED.

In equilibrium, firms that offer the highest wages survive. Hence, firms will use books if this implies that they can offer higher wages to the employees. From equations (19) and (20), the condition for using books becomes

$$1 - \tau > \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \cdot \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c}. \tag{25}$$

As long as $\tau$ is less than the Samuelson tax rate $\tau^*$, the government will be constrained by the above condition. Let $\hat{\tau}\left(A\right)$ denote the highest enforceable tax rate of the government. Then

$$1 - \hat{\tau}\left(A\right) = \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \cdot \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c}. \tag{26}$$

By log-differentiating this expression and using the envelope theorem, we obtain

$$-\frac{1}{1 - \hat{\tau}\left(A\right)}\hat{\tau}'\left(A\right) = \frac{x_A\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 0\right), A\right)} - \frac{x_A\left(N\left(A, 1\right), A\right)}{x\left(N\left(A, 1\right), A\right)}.$$

The assumption of complementarity, $x_A\left(N, A\right)/x\left(N, A\right)$ increasing in $N$, ensures that the constrained tax rate is increasing, $\hat{\tau}'\left(A\right) > 0$. The assumption (14) ensures that

$$\lim_{A \to 0} \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c} = \frac{1}{1 - c} \text{ and } \lim_{A \to \infty} \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c} = 0, \tag{27}$$

where we have used that $\lim_{A\to 0}\hat{N}(A)=0 \Rightarrow \lim_{A\to 0}N(A,1)=0$ implying that $\lim_{A\to 0}N(A,0)=0$ because of (23). In addition, we have used that $\lim_{A\to\infty}\hat{N}(A)=\infty$ implies $\lim_{A\to\infty}N(A,1)=\infty$. Thus, either $\lim_{A\to\infty}N(A,0)=$ constant in which case the assumption $\lim_{A\to\infty}\frac{x(N,A)}{x(\hat{N}(A),A)}=0$ ensures the last result or $\lim_{A\to\infty}N(A,0)=\infty$ in which case $\lim_{N\to\infty}\alpha(N)=0$ ensures the last result.

The limits in (27) and the result $\hat{\tau}'(A)>0$ imply that there exist $A_L$ and $A_H$ such that the proposition is satisfied. In particular, when $A_L\leq A\leq A_H$, the government sets $\tau(A)=\hat{\tau}(A)$ given by equation (26). QED.

## B.2 Endogenous Growth Model

### B.2.1 Households

There is a continuum (of measure one) of homogeneous individuals. Each household maximizes lifetime utility

$$u=\int_0^\infty \frac{c^{1-\sigma}-1}{1-\sigma}e^{-\rho t}dt, \tag{28}$$

where $c$ is instantaneous individual consumption (we drop time subscripts for expositional simplicity), $\rho>0$ is the rate of time preference, and $\sigma>0$ is the coefficient of relative risk aversion. We assume that each household supplies inelastically one unit of labor. The flow-budget constraint equals

$$\dot{k}=rk+(1-t)w-T-c, \tag{29}$$

where $k$ is the capital stock, $r$ is the net-return on savings, $w$ is the pre-tax labor income, $t$ is a tax rate on labor income, while $T$ is a lump sum tax. We assume that the lump sum instrument is restricted $T\leq\beta y$ where $\beta$ is the maximum fraction of aggregate income $y$ that the government can collect in lump sum taxes. We introduce lump sum taxes so that the government can raise revenue in all stages of economic development as government spending is essential for economic prosperity as we shall see below. Our empirical analysis in Figure 2 shows indeed that governments were able to raise a modest fraction of GDP in taxes before modern income and value added taxes became enforceable. Maximization of (28) subject to (29) and a no-Ponzi game condition gives the standard Keynes-Ramsey rule

$$\gamma_c\equiv\frac{\dot{c}}{c}=\frac{r-\rho}{\sigma}. \tag{30}$$

### B.2.2 Firms and Productivity

We assume that all firms have access to the same production technology and we assume that all markets are perfectly competitive. The output of firm $i$ is given by

$$y_i = f(n_i, k_i, g, k) = x\left(\frac{n_i}{\hat{n}(k)}\right) g^{\alpha} k_i^{1-\alpha} n_i^{\alpha}, \tag{31}$$

where $n_i$ is the number of employees in the firm, $k_i$ is the size of the firm's capital stock, $g$ is aggregate government spending, $k$ is the aggregate capital stock in the economy. We assume that $x(\cdot)$ is inversely U-shaped with a maximum at $n_i = \hat{n}(k)$ in which case we have $x'(1) = 0$ and $x(1) \equiv 1$. Therefore, $\hat{n}(k)$ is the optimal firm size/number of employees in the firm. Ignoring the $x(.)$ function, notice that $f(n_i, k_i, g, k)$ is homogenous of degree one in the reproducible factors of production $k$ and $g$ and is homogenous of degree one in $n_i$ and $k_i$. These two homogeneity assumptions are common in the endogenous growth literature.

Moreover, we assume that capital and firm size are complementary, $\hat{n}'(k) > 0$, reflecting that the workforce needs to organize in larger firms in order to reap the full return of a larger capital stock. Importantly, we assume for simplicity of exposition that the optimal firm's size depends on the aggregate capital stock $k$ and not on the firm's specific capital stock $k_i$. Finally, note that the capital stock of each firm is negligible compared to the aggregate capital stock as there is a continuum of firms. In a symmetric equilibrium, each firm employs $n$ workers. There is therefore a continuum of firms of measure $1/n$ (as there is a continuum of workers of measure one). Each firm employs $k_i = n \cdot k$ units of capital where $k$ is the aggregate capital stock. Hence, summing (31) across all identical $1/n$ firms, aggregate production is

$$y \equiv x\left(\frac{n}{\hat{n}(k)}\right) g^{\alpha} k^{1-\alpha} \leq g^{\alpha} k^{1-\alpha}, \tag{32}$$

which shows that aggregate output is maximized when firm size $n$ equals $\hat{n}(k)$.
Profits of firm $i$ equal

$$\pi_i = x\left(\frac{n_i}{\hat{n}(k)}\right) g^{\alpha} k_i^{1-\alpha} n_i^{\alpha} - r k_i - w n_i. \tag{33}$$

The first-order conditions with respect to $n_i$ and $k_i$ are

$$\begin{aligned} \frac{\partial \pi_i}{\partial n_i} &= x'(\cdot)\frac{1}{\hat{n}} g^{\alpha} k_i^{1-\alpha} n_i^{\alpha} + \alpha x(\cdot) g^{\alpha} k_i^{1-\alpha} n_i^{\alpha-1} - w = 0, \\ \frac{\partial \pi_i}{\partial k_i} &= (1-\alpha) x(\cdot) g^{\alpha} k_i^{-\alpha} n_i^{\alpha} - r = 0, \end{aligned}$$

which gives

$$w = \left[\frac{x'(\cdot)}{x(\cdot)}\frac{n_i}{\hat{n}} + \alpha\right] x(\cdot) g^{\alpha} k_i^{1-\alpha} n_i^{\alpha-1}, \tag{34}$$

$$r = (1-\alpha) x(\cdot) g^{\alpha} k_i^{-\alpha} n_i^{\alpha}. \tag{35}$$

From equations (33)–(35), we obtain

$$\pi_i = -\frac{x'(\cdot)\frac{n_i}{\hat{n}}}{x(\cdot)} x(\cdot) g^{\alpha} k_i^{1-\alpha} n_i^{\alpha}$$

Free entry and exit ensures that profits are zero therefore entry/exit will occur until $n_i = \hat{n}(k)$ such that $x'(1) = 0$.

The aggregate production, wage rate, and real interest rate become

$$y = g^{\alpha} k^{1-\alpha}, \tag{36}$$

$$w = \alpha g^{\alpha} k^{1-\alpha}, \tag{37}$$

$$r = (1-\alpha) g^{\alpha} k^{-\alpha} \tag{38}$$

where we have used equations (32), (34), (35), and $n_i = \hat{n}(k)$. Note that the standard macroeconomic equation $y = w + rk$ holds.

### B.2.3 Optimal Government Policy

*Case with No Tax Evasion*

We consider a benevolent government that chooses $(g, T, t)$ in order to maximize lifetime utility (28). The government policy has to satisfy the government budget constraint

$$g \leq T + tw. \tag{39}$$

Let $\tau \equiv g/y$ denote the government to GDP ratio. From equation (36), we have

$$g/k = \tau^{\frac{1}{1-\alpha}} \tag{40}$$

From equations (29), and (39) and (40), we obtain the resource constraint

$$\dot{k} = g^{\alpha} k^{1-\alpha} - g - c = \tau^{\frac{\alpha}{1-\alpha}} k - g - c. \tag{41}$$

From equations (30) and (33), we obtain the steady state growth rate of consumption for a given government to GDP ratio $\tau$

$$\gamma_c = \frac{(1-\alpha)(g/k)^{\alpha} - \rho}{\sigma} = \frac{(1-\alpha)\tau^{\frac{\alpha}{1-\alpha}} - \rho}{\sigma}, \tag{42}$$

which also becomes the steady state growth rate of $k$ and $y$. The marginal benefit of raising public spending is $\alpha g^{\alpha-1}k^{1-\alpha} = \alpha\,(g/k)^{\alpha-1}$ while the marginal cost is 1. This, together with equation (40), implies that the optimal policy solution that decentralizes the first best allocation is $\tau^* = \alpha$ in which case, the growth rate of the economy becomes

$$\gamma_c = \frac{(1-\alpha)\,\alpha^{\frac{\alpha}{1-\alpha}} - \rho}{\sigma}, \tag{43}$$

which is constant over time.[43]

*Case with Full Tax Evasion*

With full tax evasion, it is impossible to tax income, $t = 0$. We now have$g = T \leq \beta y$. We assume $\beta < \alpha$ implying that it is impossible to attain the optimal government-to-GDP ratio with lump sum taxation alone. From (42), we obtain the growth rate

$$\gamma_c = \frac{(1-\alpha)\,\beta^{\frac{\alpha}{1-\alpha}} - \rho}{\sigma}. \tag{44}$$

The growth rate will be positive or negative depending on whether $\beta$ is above or below $\left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$.

*Case with Tax Enforcement*

We consider the whistleblower model of tax evasion. We denote by $\bar{y}$ the net-of-tax income of each employee in his best outside option, where $\bar{y}$ is determined by the equilibrium in the labor market and taken as given by the firm. The firm then has to offer each employee a pre-tax compensation equal to $\bar{y}/(1-t)$ if it complies with the tax law, and equal to $\bar{y}$ if it evades all taxes.

If the firm evades and nobody whistleblows, the income of each employee is given by $w = \bar{y}$. If an employee decides to whistleblow (given that nobody else does), he can obtain income $\bar{y} - \tau(1+\theta)\bar{y} + \delta(1+\theta)\tau\bar{y}n_i$. The employee is therefore prevented from whistleblowing iff $\bar{y} \geq \bar{y} - \tau(1+\theta)\bar{y} + \delta(1+\theta)\tau\bar{y}n_i$, which is equivalent to $n_i \leq 1/\delta$ as in Proposition 4. Hence, a firm that evades tax has to choose a firm size $n_i$ below $1/\delta$. We can prove the following Proposition:

**Proposition 11** *Let $\bar{n} \equiv 1/\delta$. We obtain the following cases:*

*(1) If $\hat{n}(k) \leq \bar{n}$, then the firm evades all taxes and chooses the optimal firm size $\hat{n}(k)$.*

[43]We assume $(1-\alpha)\,\alpha^{\frac{\alpha}{1-\alpha}} > \rho > (1-\alpha)\,\alpha^{\frac{\alpha}{1-\alpha}}(1-\sigma)$, where the first inequality ensures a positive growth rate while the second inequality ensures that utility is bounded.

*(2) If $\hat{n}(k) > \bar{n}$ then:*

*(a) If $t > 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}$, the firm evades all taxes and chooses suboptimal firm size $\bar{n}$.*

*(b) If $t \leq 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}$, the firm does not evade and chooses the optimal firm size $\hat{n}(k)$.*

**Proof:** Without tax evasion, $n_i = \hat{n}(k)$ is optimal and the pre-tax wage rate as a function of the capital stock is given by equation (37) such that $w = \alpha \cdot g \cdot (k/g)^{1-\alpha}$. The capital stock as a function of the real interest rate from equation (38) is such that $k = \left(\frac{1-\alpha}{r}\right)^{\frac{1}{\alpha}} g$. By inserting this expression, in equation (37), we obtain

$$\bar{y} = (1-t)\, w = (1-t)\, \alpha g \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}}. \tag{45}$$

Taxation is sustainable if a single firm cannot achieve a higher profit by cheating. Since, profit is zero in the no-evasion equilibrium, this requirement implies that profits with tax evasion are negative:

$$\pi_i^E = x\left(\frac{n_i}{\hat{n}(k)}\right) g^{\alpha} k_i^{1-\alpha} n_i^{\alpha} - r k_i - \bar{y} n_i \leq 0.$$

If $\hat{n}(k) > \bar{n}$, then the optimal size choice for the evading firm is $n_i = \bar{n}$. In that case, the optimal capital stock if the firm evades becomes

$$\frac{\partial \pi_i^E}{\partial k_i} = (1-\alpha)\, x(\cdot)\, g^{\alpha} k_i^{-\alpha} \bar{n}^{\alpha} - r = 0.$$

By isolating $k_i$ and substituting the result back into the profit expression, we arrive at the condition

$$\pi_i^E = \left[\alpha x(\cdot)^{\frac{1}{\alpha}}\, g \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}} - \bar{y}\right] \bar{n} \leq 0.$$

The company has to offer each worker at least $\bar{y}$ in (45). This implies

$$\pi_i^E = \left[x(\cdot)^{\frac{1}{\alpha}} - (1-t)\right] \alpha g \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}} \bar{n} \leq 0,$$

which is fulfilled if

$$t \leq 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}. \tag{46}$$

Using the same procedure, it is possible to show the reverse result, i.e., starting from an evasion equilibrium, it is not possible to obtain a higher profit by not evading if $t > 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}$.

The proof of (1) follows from the fact that profits are always greater under evasion when this can be sustained at the optimal firm size $\hat{n}(k)$. The proof of (2a) and (2b) follows directly from the above argument and the condition (46). QED.

*Macroeconomic Development*

We now characterize the optimal government policy and the macroeconomic development of the economy. Let us denote by $\underline{k}$ the aggregate capital stock that solves $\hat{n}(\underline{k}) = \bar{n} \equiv 1/\delta$, and let $\bar{k}$ be the capital stock that solves $x\left(\bar{n}/\hat{n}\left(\bar{k}\right)\right)^{\frac{1}{\alpha}} = \beta/\alpha < 1$. As $\hat{n}\left(\bar{k}\right) > \bar{n}$, we have $\underline{k} < \bar{k}$. We consider an economy with an initial capital stock below $\underline{k}$. We have

**Proposition 12** *Optimal government policy and possible stages of economic development*

*(1) Poverty trap: If* $\beta \leq \left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$ *then the government cannot raise income taxes and the economy will experience negative growth.*

*(2) Economic development: If* $\beta > \left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$ *then:*

*(a) First stage (underdeveloped economy): When* $k < \underline{k}$*, the government cannot raise any tax revenue. The growth rate of the economy is positive but too low compared to the first-best growth rate.*

*(b) Intermediate stage: When* $\underline{k} \leq k \leq \bar{k}$ *, the government is constrained by tax enforcement and sets* $t = 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}$*. Firms do not evade taxes but government size is suboptimal* ($\tau^* < \alpha$)*. The growth rate of the economy is positive and increasing but too low compared to the first-best growth rate.*

*(c) Last stage (modern economy): When* $k > \bar{k}$ *, the government is no longer constrained by tax enforcement, firms do not evade taxes, government size is socially optimal* ($\tau^* = \alpha$)*, and the growth rate of the economy equals the the first-best growth rate.*

**Proof:** In all cases, the economy starts with $k < \underline{k}$ so that firms are untaxable initially. Suppose that $\beta \leq \left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$, then equation (44) implies that the growth rate is negative. In that case, the economy is stuck in a poverty trap which proves (1).

Suppose instead that $\beta > \left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$, then equation (44) implies that the growth rate is positive. As $\beta < \alpha$, the growth rate is lower than the first best growth rate given by (43) which proves (2a).

As the economy has a positive growth rate, $k$ will eventually reach $\underline{k}$ and Proposition 11, (2b) implies that a maximum tax at rate $t = 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}$ can be enforced, which proves (2b).

Eventually, $k$ will reach $\bar{k}$ at which point the first best tax rate $\tau^* = \alpha$ can be enforced and the growth rate becomes first best optimal which proves (2c). QED.

# References

**Acemoglu, Daron, Robinson, J.A.** (2000) "Why did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics* 115(4), 1167–1199.

**Aidt, Toke S. and Peter S. Jensen** (2009) "Tax Structure, Size of Government, and the Extension of the Voting Franchise in Western Europe, 1860-1938", *International Tax and Public Finance* 16(3), 362-394.

**Allingham, M. G. and A. Sandmo** (1972). "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1, 323-338.

**Alt, J.** (1983), "The Evolution of Tax Structures", *Public Choice* 41(1), 181-222.

**Andreoni, J., B. Erard, and J. Feinstein** (1998). "Tax Compliance." *Journal of Economic Literature* 36, 818-860.

**Azariadis, Costas and John Stachurski** (2005). "Poverty Traps," in Philippe Aghion and Steven Durlauf (eds.), *Handbook of Economic Growth* Vol. 1A, North-Holland, Amsterdam.

**Barro, Robert J.** (1990) "Government Spending in a Simple Model of Endogenous Growth," *Journal of Political Economy* 98(5), Part 2, S103-S125.

**Baumol, W.J.** (1967), "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *American Economic Review* 57(3), 415-426.

**Baumol, W.J. and W.G. Bowen** (1966), *Performing Arts: The Economic Dilemma*, Twentieth Century Fund: New York.

**Bayer, R.-C. and F. A. Cowell** (2005). "Tax compliance and firms' strategic interdependence," Distributional Analysis Discussion Paper 81, STICERD, London School of Economics, revised September 2008.

**Becker Gary S.** (1968), "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76(2), 169-217.

**Besley, Timothy and Torsten Persson** (2008), "The Origins of State Capacity: Property Rights, Taxation and Politics", NBER working paper No. 13028, forthcoming in the *American Economic Review*.

**Besley, Timothy and Torsten Persson** (2009) "State Capacity, Conflict and Development", LSE Working Paper, June 2009.

**Bird, Richard M.** (1989). "The Administrative Dimensions of Tax Reform in Developing

Countries." In *Tax Reform in Developing Countries*, edited by Malcom Gillis. Durham: Duke University Press.

**Bird, R. M.** (1992), *Tax Policy and Economic Development*, John Hopkins University Press.

**Bird, Richard M.** (2003), "Administrative Dimensions of Tax Reform" ITP Paper 0302.

**Bird, Richard M. and Oliver Oldman** (1990). *Readings on Taxation in Developing Countries*, John Hopkins University Press, 4th Edition.

**Brennan, G. and J.M. Buchanan** (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution*, Cambridge University Press, Cambridge.

**Chen, K.-P. and C. Y. C. Chu** (2005). "Internal Control versus External Manipulation: A Model of Corporate Income Tax Evasion," *RAND Journal of Economics* 36, 151-164.

**Cowell, Frank.** (1990) *Cheating the Government*. MIT Press: Cambridge MA.

**Crocker, Keith and Joel Slemrod.** (2005) "Corporate Tax Evasion with Agency Costs," *Journal of Public Economics* 89, 1593-610.

**Dharmapala, Dhammika, Slemrod, Joel, Wilson, Jay.** (2008) "Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs," University of Michigan Working Paper.

**Dixit, Avinash K.** (2004) *Lawlessness and Economics: Alternative Modes of Governance,* Princeton University Press: Princeton.

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

**Eurostat.** (2007) "Undeclared Work in the European Union," *Special Eurobarometer* 284 from the European Commission.

**Flora, Peter** (1983) *State, Economy, and Society in Western Europe, 1815-1975*, Volume I, Macmillan Press: London.

**Gillis, M.** (1989) *Tax Reform in Developing Countries.* Durham: Duke University Press.

**Gordon, Roger and Wei Li** (2005) "Tax Structure in Developing Countries: Many Puzzles and a Possible Explanation," NBER Working Paper No. 11661, forthcoming *Journal of Public Economics*.

**Hesch, Joel D.** (2002) *Reward: Collect Millions for Reporting Tax Evasion (A Guide to the IRS Whistleblower Reward Program)*, LU Books.

**Hettich, W., and Winer, S.** (1991). "Debt and Tariffs: An Empirical Investigation of the

Evolution of Revenue Systems," *Journal of Public Economics* 45, 215–242.

**Hinrichs, H.H.** (1966), *A General Theory of Tax Structure Change During Economic Development,* (The Law School of Harvard University, Cambridge, MA).

**Internal Revenue Service.** (2006) *Tax Year 2001 Federal Tax Gap*, IR-2006-28, Washington, D.C.

**Kau, James B. and Paul H. Rubin.** (1981) "The Size of Government", *Public Choice* 37(2), 261-274.

**Keen, Michael.** (2007) "VAT, Tariffs, and Withholding: Border Taxes and Informality in Developing Countries", IMF Working Paper WP/07/174.

**Kelley, Patrick L. and Oliver Oldman.** (1973) *Readings on Income Tax Administration*, The Foundation Press, New York.

**Kenny, L.W., and Winer, S.** (2006). "Tax Systems in the World: An Empirical Investigation into the Importance of Tax Bases, Administrative Costs, Scale and Political Regime," *International Tax and Public Finance* 13, 181–215.

**Kleven, Henrik, Martin Knudsen, Claus Kreiner, Soren Pedersen, and Emmanuel Saez.** (2009) "An Experimental Evaluation of Tax Evasion and Tax Enforcement in Denmark", preliminary draft August 2009.

**Kopczuk, Wojciech, and Joel Slemrod** (2006). "Putting Firms into Optimal Tax Theory", *American Economic Review: Papers and Proceedings* 96(2), 130-134.

**Lederman, Leandra** (2009) "Reducing Information Gaps to Reduce the Tax Gap: When Is Information Reporting Warranted?", Indiana University Legal Studies Research Paper No. 126.

**Lindert, Peter H.** (2004). *Growing Public: Social Spending and Economic Growth since the Eighteenth Century* (Volumes 1 and 2). Cambridge: Cambridge University Press.

**Logue, Kyle D. and Joel Slemrod** (2008) "The Coase Theorem and Tax Law", Working Paper 2008.

**Moore, John** (1992). "Implementation, Contracts, and Renegotiation in Environments with Complete Information," in J-J. Laffont (ed.), *Advances in Economic Theory: Invited Papers for the Sixth World Congress of the Econometric Society*, 182-281.

**Musgrave, Richard** (1966) *Fiscal Systems*, Yale University Press: New Haven.

**OECD** (2004). *Employment Outlook*, Chapter 5, "Informal Employment and Promoting the Transition to a Salaried Economy", Organization for Economic Co-operation and Development:

Paris, 225-289.

**OECD** (2006). *Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series,* Organization for Economic Co-operation and Development: Paris.

**OECD** (2008). *Revenue Statistics: 1965-2007,* Organization for Economic Co-operation and Development: Paris.

**Peacock, Alan and Jack Wiseman** (1961) *The Growth of Public Expenditure in the United Kingdom*, NBER: New York.

**Peltzman, Samuel** (1980) "The Growth of Government" *Journal of Law and Economics* 23(2), 209-287.

**Riezman, R., and Slemrod, J.** (1987). "Tariffs and Collections Costs," *Review of World Economics* 123, 545–549.

**Schneider, Friedrich and Dominik H. Enste** (2000) "Shadow Economies: Size, Causes and Consequences," *Journal of Economic Literature* 38(1), 77-114.

**Shaw, Jonathan, Slemrod, Joel, and John Whiting** (2008) "Administration and Compliance", IFS Working Paper, forthcoming in *Reforming the Tax System for the 21st Century: The Mirrlees Review.*

**Slemrod, Joel** (2004) "The Economics of Corporate Tax Selfishness" *National Tax Journal* December, 57(4), 877–99.

**Slemrod, Joel** (2007) "Cheating Ourselves: The Economics of Tax Evasion," *Journal of Economic Perspectives* 21(1), Winter, 2007, 25-48.

**Slemrod, Joel, and Shlomo Yitzhaki** (2002) "Tax Avoidance, Evasion and Administration," in: A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics* Vol. 3, North-Holland, Amsterdam, 1423-1470.

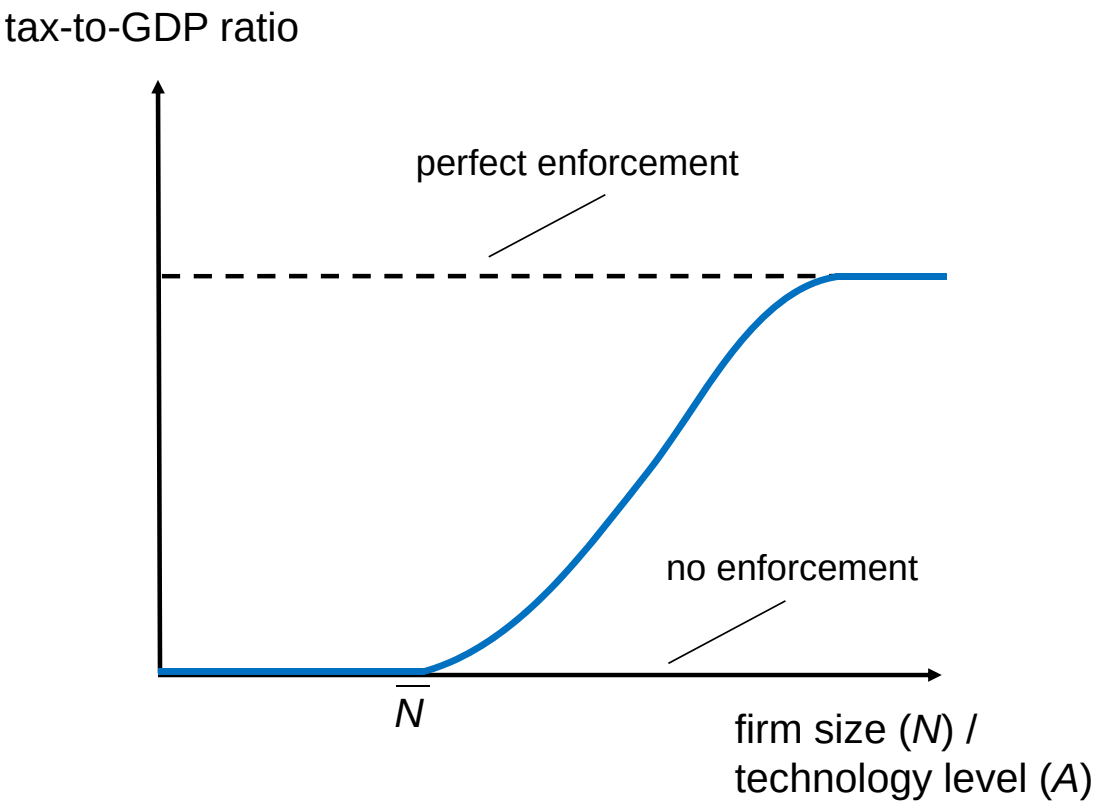
**Surrey, Stanley S.** (1958) "Tax Administration in Underdeveloped Countries", University of Miami Law Review, 12, 158-188.

**Tybout, James R.** (2000). "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature* 38(1), 11-44.

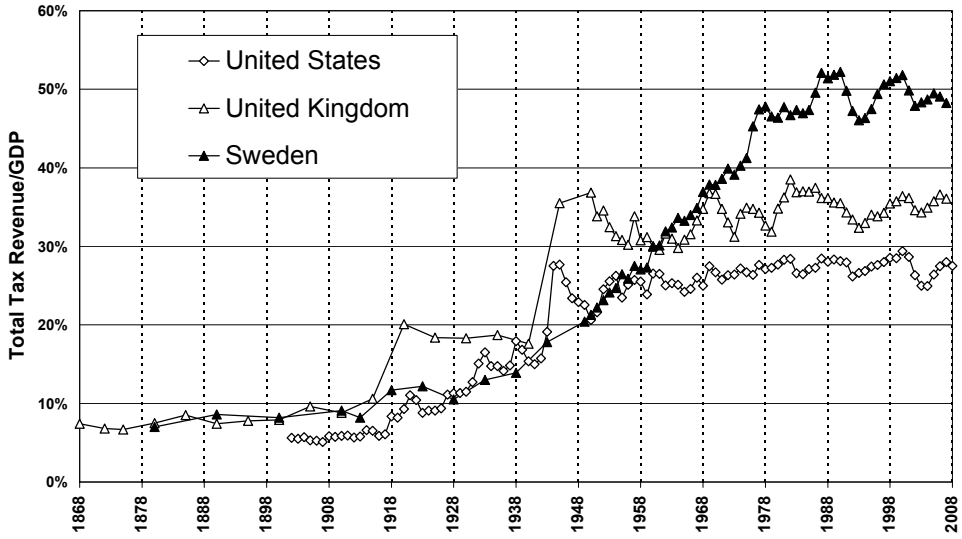
**Webber, Carolyn, and Aaron B. Wildavsky** (1986). *A History of Taxation and Expenditure in the Western World.* New York: Simon and Schuster.

**Yaniv, Gideon** (1992) "Collaborated Employee-Employer Tax Evasion," *Public Finance* 47(2), 312-321.

Figure 1
Evolution of Government Size



**A. Tax revenue to GDP ratio in the US, UK, and Sweden**




**B. Evolution of income vs. other taxes in the US, 1902-2008**

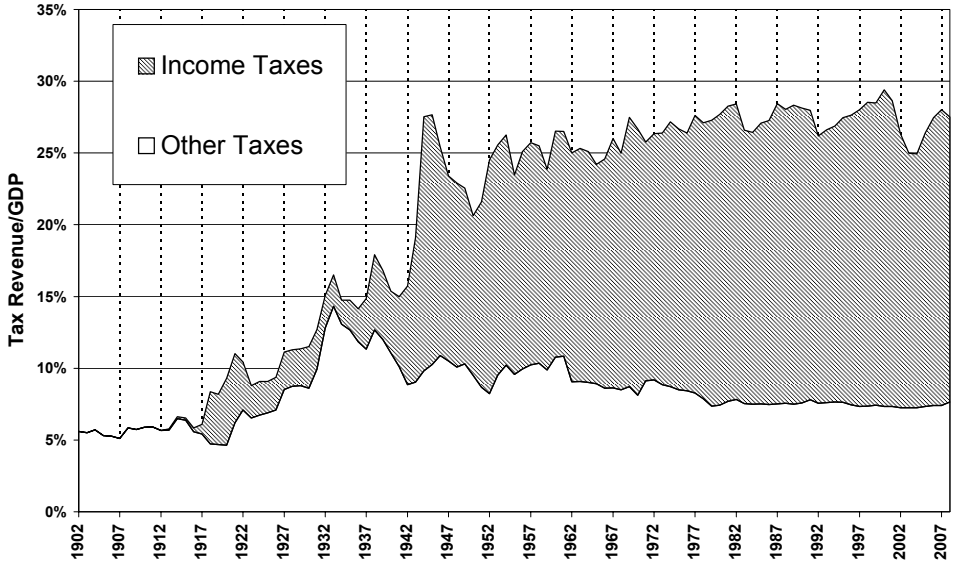



**FIGURE 2**
**Evolution of Tax Revenue to GDP in Three Advanced Economies**

Sources: United States: 1902-1995 from the Historical Statistics of the United States, 1996-present from the annual Statistical Abstracts of the United States. Series have been compiled with interpolations for some missing years in the period 1902-1949 by Christopher Chantrill at http://www.usgovernmentrevenue.com/

United Kingdom and Sweden: 1868-1964 from Flora (1983). 1965-present from OECD, Tax Revenue Statistics and GDP Statistics.

Notes: In panel A, taxes include taxes from all levels of government (central, state, and local). Taxes include social security contributions. We have used series from Flora et al. for general tax revenue (as a percent) of GDP and adding social security receipts paid by employees and employers. Period 1965-present, we use total tax revenue from OECD tax statistics, GDP series are from OECD, National Income Statistics. In panel B, income taxes include the individual and corporate federal and state income taxes, all payroll taxes financing social insurance programs. Other taxes include all other taxes (primarily property taxes, sales taxes, excise taxes, custom duties, estate and inheritance taxes, and various other small taxes).