Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries*

Henrik Jacobsen Kleven, London School of Economics and CEPR
Claus Thustrup Kreiner, University of Copenhagen and CESifo
Emmanuel Saez, UC Berkeley and NBER

Working Draft: May 2009

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 accounting books. Because such books are widely used within the firm, any single employee can denounce collusive tax cheating between employees and the employer by revealing the true books 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 problems, 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. Therefore, capitalistic development relaxes the tax enforcement constraint and naturally leads to large welfare state governments.

*We would like to thank David Albouy, Saki Bigio, Raj Chetty, Wojciech Kopczuk, Camille Landais, Joel Slemrod, and numerous seminar participants for helpful discussions and comments. Corresponding author: Emmanuel Saez, University of California, Department of Economics, 549 Evans Hall #3880, Berkeley, CA 94720, email: saez@econ.berkeley.edu. Financial support from NSF Grant SES-0850631 is gratefully acknowledged.

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 \ rms") generally have a large number of employees or clients, and therefore need to use accurate and rigorous accounting books to carry out their business activities. Firms report taxable income | paid to employees or clients | directly to the government, and therefore act as a third party between taxpayers 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 2002, 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 rst 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. Similarly, Kleven et al. (2009) study an income tax audit experiment in Denmark and nd that, although purely self-reported income constitutes only about 8% of total reported income, it accounts for about 90% of detected evasion.

In spite of its central importance, the theoretical literature on tax evasion and tax enforcement 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 in practice and which covers a minor part of taxation in advanced

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

economies.[2] 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?[3]

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 and paid by firms (middle tier). The firm acts as a third party that reports income on behalf of the employees. 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 company such as a bank, a mutual fund, or a pension fund. When a firm is large and complex, using accounting books is extremely valuable for productivity. Such accounting books are widely used within the firm and hence many employees know about them. Such books record accurately the amounts paid to each employee as wages and salaries.

In principle, the firm and its employees could collude to report smaller incomes to the government than those actually paid out by the firm. 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 books 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 accounting books 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 hard book evidence which makes third-party tax enforcement so successful.

Firms trade off the value of using accounting books (for productivity) against the risk of

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

[3] A large number of studies suggest that the observation of 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); Frey 1997; Andreoni et al. 1998). In this paper, we propose instead a theory that focuses on information reasons for high compliance.

being caught by the government when colluding with employees to evade taxes. When firms are large, the optimal choice is to have accurate books and not cheat at all. The size threshold above which firms use accurate books is larger when taxes are higher and smaller when books are more important to the business Our results are consistent with a body of work on tax enforcement in developing countries showing that tax evasion (the \informal sector") is concentrated among small firms (see, e.g., Levy 2007 for a recent study of the case of Mexico).

We embed this agency model into the standard 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 taxfilers who adjust self-reported income correspondingly. As a result, the level of tax evasion ends up being independent of third-party reporting. 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, governments often disallow self-reported income losses to count against other income items. Tax administrations also concentrate audits 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. We argue that those findings help explain the historical development of income taxation.

The last part of the paper embeds our agency model into a simple macroeconomic growth model where the size of firms grows with exogenous technological progress. In the model, a representative individual has preferences over private and public goods (broadly defined to represent all forms of goods potentially supplied by the government). For simplicity, we ignore supply-side responses to taxation such as labor supply and savings responses. As a result, 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, 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. 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 for their own consumption under a fiscal capacity constraint.[4]

This simple macro-model can account for the historical growth in government size over the 19th and 20th century and the stability of government size since the 1970s in the most advanced 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 need to use accounting books and can then be easily taxed by the government. 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 related literature. Section 3 presents our agency model of third-party tax enforcement. Section 4 embeds this model into a standard Allingham-Sandmo model of tax evasion. Section 5 embeds the micro-model in a simple macroeconomic framework, which accounts for the evolution of government size over the course of development. Finally, Section 6 offers concluding remarks and plans for future empirical tests of our theory.

[4]While it follows straightforwardly that the first two stages of development (where imperfect enforcement constrains government growth) is consistent with a Leviathan view of government, the last stage of development where enforcement is perfect and government size is stable could be consistent with Leviathan if we incorporate real behavioral responses to taxation into the model. In this case, the government size to output would ultimately converge to the tax rate maximizing tax revenue.

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 (1988, 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 veri able books and many employees.

Recently, a number of studies have made important progress in modeling the problem of tax enforcement. First, Gordon and Li (2005) develop a model where the government can collect taxes only from formal rms de ned as those connected to the nancial sector. They show how the lessons from optimal tax theory are drastically changed in this environment and t 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 modeling the tax evasion game in a three-tiered agency model. Second, Kopczuk and Slemrod (2006) set out a simple model to show how the network of rm-to- rm 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 rm-to- rm transaction information. Our paper focuses on the within- rm information network rather than the across- rm information network and is therefore complementary to the Kopczuk-Slemrod analysis. Third, a number of studies in the corporate income tax evasion literature have shown that the internal organization or the external activities of rms can a ect their tax reporting decisions. 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 e ective in reducing evasion than penalties imposed on shareholders. Chen and Chu (2005) show that the evasion decision of the rm's owner a ects the optimal compensation scheme o ered to employees and hence creates a distortion in the manager's e ort and reduces the e ciency of the contract. On the external side, Bayer and Cowell (2005) show that imperfect competition between rms have important consequences for the e ciency e ects of corporate tax audits.

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 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, 1996). Third, Peacock and Wiseman (1961) proposed a \ratchet e ect 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 a ected by the temporary shock. Notice that the Wagner, Baumol, and ratchet e ect 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, who are unchecked by electoral constraints (Brennan and Buchanan, 1980). Governments therefore act as monopolies, which maximize revenue under constitutional and scal 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, legislative bargaining, 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, 1994, 2004).

In addition to these hypotheses, a number of studies have pointed out that there are scal capacity constraints to government growth (e.g., Kau and Rubin, 1981; Bird, 1983, 1989, 1992; Peltzman, 1980; Riezman and Slemrod, 1987; Becker and Mulligan, 2003; Kenny and Winer, 2006; Dusek, 2006; Aidt and Jensen, 2008). Moreover, there is a vast literature on the role of

under-development in constraining tax structures both historically and in current developing countries.[5] Our theory proposes a micro-foundation that accounts for the changes in scal constraints over the course of development.

Recently, Besley and Persson (2007, 2008) propose an extension of the Ratchet e ect theory that emphasizes the increasing scal capacity over the course of development. They develop a model where governments invest in scal 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 scal 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 5). 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-Coate papers. Our theory explains the increasing scal capacity and government size over the process of development by increasing rm size/complexity and the associated use of rigorous accounting books, which ultimately makes third-party tax enforcement e ective.

3 A Microeconomic Theory of Third-Party Reporting

3.1 Firm-Employee Game with Accounting Books

3.1.1 Model

We start by analyzing the case of a rm using accounting books, and study the decision to use books in Section 3.2. The rm has N employees and generates a given total revenue W. Employees are paid wages $w = (w_1, ..., w_N)$ and pro ts are zero, so that $\sum_{n=1}^{N} w_n = W$.[6] The rm and employees agree on reports to the government given by $w = (w_1, ..., w_N)$, 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, w) are determined cooperatively by the N

[5] See, e.g., Alt (1983), Avi-Yonah (2000), Bird and Oldman (1975, 1990), Gillis (1989), Hettich and Winer (1991), Hinrichs (1966), Kelly and Oldman (1973), Kenny and Winer (2006), Webber and Wildavsky (1996).

[6] There would be no substantial change in our results if we assumed that one of the employees were instead the employer and that his compensation were equal to the profits of the firm. We discuss this issue in more detail in Section 3.3.

employees of the firm. Because this is a tax collusion game, a cooperative game seems to be the most natural one.[7] 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 $y = (y_1, ..., y_N)$ the disposable income levels (net of taxes) associated with those outside options.[8] In the general equilibrium macro-model presented in Section 5, we fully endogeneize outside options and firm size N.

The presence of books creates common knowledge within the firm. We capture such common knowledge by assuming that (w, w) is known to everyone within the firm. In practice, although books 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 $w_n \neq w_n$ are aware of tax evasion and can denounce tax cheating within the firm.

Let us assume further that books create verifiable information in the following sense: If any employee reveals the book information (w, 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, w) with the cooperation of the denouncing employee. Because true books are widely used within the company, it is impossible to hide them if a single knowledgeable insider is determined to reveal the true book 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 books. For simplicity, in that case, we assume that the government cannot detect cheating at all.

Following the report w to the government, taxes are paid at rate τ based on w. Each employee $n = 1, ..., N$ then decides either to stick to the report w_n or to whistleblow and reveal the true books to the government if $w \neq w$. If nobody whistleblows, the government cannot effectively audit the firm, and tax evasion, if any, will go undetected. On the other hand, if

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

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

any employee whistleblows and brings the true book information to the government, then the government will audit the rm and will be able to fully detect tax evasion. When evasion is detected, we assume that the government charges the evaded tax plus a penalty equal to θ percent of the evaded tax to each person caught evading. In addition, the government may o er a reward to whistleblowers equal to a share δ of total uncovered tax evasion. For simplicity, we assume that all workers are risk neutral.[9]

We may summarize the timing of the game as follows: (1) employees agree cooperatively on a vector of wages $(w_1, ..., w_N)$ and a vector of reports $w = (w_1, ..., w_N)$, (2) taxes are paid based on w at rate τ, (3) each employee n decides to stick to the report w_n or to whistleblow if $w \neq w$, and (4) the government decides to audit or not, and nes and potential whistleblower rewards are paid.

For simplicity of exposition, we have modelled the rm as a passive player that makes no pro ts. In this context, for each employee n, we may think of the \third party" as being represented by all the co-workers, who may or may not report a discrepancy between true and reported wages of the employee. The \ rm" is simply a group of workers entering into a production process and sharing information so that each employee knows (w, w) and can either collude or whistleblow.

In this model, if all employees can commit ex ante never to denounce tax cheating to the government, then the collusive equilibrium where $w = 0$ and no taxes are paid is sustainable and is the optimal solution from the perspective of the rm. In other words, and as pointed out by Yaniv (1988, 1992), under perfect collaboration between employees and employers, the employees-employer forms a single unit and the model reverts to the standard Allingham-Sandmo model where a single taxpayer plays a tax evasion game with the government.

However, the perfect commitment assumption is unlikely to hold in practice. There are two sets of reasons why employees may denounce tax cheating to the government. The rst set of reasons is the presence of $random\ shocks$ such as a con ict between an employee and the employer, moral concerns of a newly hired employee about defrauding the government, or simply a mistake whereby an employee reveals the true books w to the government instead of the fake books w. The second reason is the presence of rational whistleblowing if the government o ers

[9] Assuming risk aversion would make tax enforcement easier for the government. We consider risk aversion in Section 3 in the context of the Allingham-Sandmo model.

a reward to the whistleblower. We develop both models below and show that, when firms are large, tax evasion is bound to be uncovered which deters it in the first place.

3.1.2 Random Shock Model

We incorporate the possibility that an employee may deviate and reveal books either by mistake, because he is disgruntled, or because of moral concerns.[10] Let ε be the probability of any given employee revealing true books 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 books (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 pay-off of each employee equals

$$y_n = w_n - \tau \cdot w_n - (1-(1-\varepsilon)^N) \cdot \tau \cdot (1+\theta) \cdot (w_n - w_n)^+.$$

We assume that workers decide cooperatively on vectors of true and reported wages (w, 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, 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 $w_n, w_n \geq 0$ for all n, and participation constraints $y_n \geq y_n$ for all n, ensuring that each employee obtains a payoff that is at least as high as his best available outside option y_n. The coalition of workers $1, ..., N$ will find it optimal to increase or decrease the report w_n for worker n depending on the derivative of total surplus with respect to w_n. When $w_n < w_n$, we have:

$$\frac{\partial Y}{\partial w_n} = \tau \cdot [-1 + (1+\theta)(1-(1-\varepsilon)^N)]. \quad (1)$$

When $w_n > w_n$, we have: $\frac{\partial Y}{\partial \bar{w}_n} = -\tau$, so that it never pays to over-report wages.[11]

Proposition 1 *In the random shock model, any cooperative solution is such that:*
(a) If $(1-\varepsilon)^N \leq \theta/(1+\theta)$, *there is no tax evasion at all:* $w = w$.

[10] 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.

[11] 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 $\bar{w}_n > w_n$.

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

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

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

Four points are worth noting about Proposition 1. First, when $\varepsilon = 0$, we are back to the standard case analyzed by Yaniv (1988, 1992) where rm size does not help and there is always tax evasion.[12] Second, when $\varepsilon > 0$ and even for moderate nes $\theta > 0$, it will always be the case that large rms choose not to evade and hence there will be no need for high audit rates. Our model can therefore explain why low nes 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 nes are low (given reasonable risk aversion parameters). The key feature of the model that generates successful tax enforcement is the inability of workers to commit ex-ante to not denounce cheating to the government. 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 speci c division of revenue $W = w_1 + ... + w_N$ across workers. The equilibrium division will depend on the outside opportunities y and other factors not explicitly speci ed 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, w. An alternative polar assumption is that each worker knows only about his/her own wages w_n, w_n. This situation is more realistic in the case of a nancial institution such as a mutual fund handling the accounts of many clients and reporting the returns of those accounts to clients and the government. Critically, we maintain the assumption that, if there is under-reporting for individual n ($w_n < w_n$) and individual n denounces the rm, the government will carry out an audit and then be able to observe the full set of actual and reported wages w, w. This

[12] Tax evasion is not complete in Yaniv due to risk aversion as in the standard Allingham-Sandmo model.

assumption can be defended as follows. A formal business needs to record w and w in the books. Individual n can prove that $w_n \neq w_n$ as long as w_n was formally paid out. Therefore, with hard evidence that the rm cheated on individual n, an investigation may be able to retrieve the true books and obtain full information w, w. In other words, the rm is a nexus of information written in the accounting books, and the information cannot be broken or hidden into isolated pieces.

Proposition 2 *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 the largest integer below 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 rm. Hence, the total surplus is given by

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

When $w_n < w_n$, we have:

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

Therefore, evasion is pro table only if $N_c \leq N$ de ned as $[1 - (1 - \varepsilon)^{\bar{N}}](1 + \theta) = 1$. When an employee evades, the surplus is maximized by full evasion: $w_n = 0$. Because the extra surplus created by full evasion is proportional to w_n, surplus is maximized by having the highest-paid employees evade.
(b) Because N_c is xed, as N goes to in nity, a vanishing fraction of employees will be able to evade. If the wage distribution is xed, the share of total compensation going to a vanishing fraction of employees also converges to zero. QED.

Two points are worth noting about Proposition 2. First, our results of successful enforcement for large rms 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 rms. 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 will therefore limit the flow of information, and discard the use of accounting books in the most extreme cases (see Section 3.2).

3.1.3 Rational Whistleblower Model

We now consider the case where the government offers a whistleblower reward equal to a fraction δ of total under-reported tax liability. Hence, employees have a monetary incentive to denounce tax cheating. Many OECD countries run 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, 2009). 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 δ dollars.

In principle, the firm and the employees could sustain the collusive tax cheating equilibrium if employees could commit ex ante not to whistleblow. Our key assumption is that such ex ante commitment is not feasible. In practice, firms do not have the power to enforce non-whistleblowing commitments.[13]

Given payments $w = (w_1, ..., w_N)$ and reports $w = (w_1, ..., w_N)$, the payoff for employee n if he does not whistleblow is given by

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

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

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

[13] 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. In a dynamic model, it is conceivable that whistleblowers could be fired and hence loose future rents from the employment match. Such an extension would make enforcement harder, but would not change the essence of our results.

where N_w denotes the number of whistleblowers who share equally the rewards from whistleblowing.

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

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

A cooperative solution (w, w) maximizes surplus Y subject to $\sum_{n'} w_{n'} = W$, non-negativity constraints $w_n, w_n \geq 0$ for all n, and participation constraints $y_n \geq y_n$ for all n. Notice that $\theta > \delta$ 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,

$$\delta \leq \frac{(1 + \theta)\left(w_n - w_n \right)^{+}}{\sum_{n'} \left(w_{n'} - w_{n'} \right)^{+}}. \tag{5}$$

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

Proposition 3 *In the whistleblower model, any cooperative solution is such that:*
(a) If $N > (1 + \theta)/\delta$, then there can be no tax evasion at all: $w = w$. Hence large firms do not evade taxes even if $\delta > 0$ is very small.
(b) If $N \leq (1 + \theta)/\delta$, 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 $N > (1 + \theta)/\delta$ and that there is some evasion $E \equiv \sum_{n'} (w_{n'} - w_{n'}) > 0$. Then, from eq. (5), we have $(1 + \theta)(w_n - w_n) \geq \delta E$ for all n. Summing across all

n, this implies (1 + θ)$E \geq \delta \cdot N \cdot E$. Because $E > 0$, this implies 1 + $\theta \geq \delta \cdot N$, which is a contradiction.

For (b), if some evasion is sustained ($E > 0$), then we must have $w_n - w_n \geq \frac{\delta}{1+\theta} E$ for all n. Because $\frac{\delta}{1+\theta} \leq \frac{1}{N}$ in this case, it is feasible to satisfy this condition, for example by having equal evasion across all employees: $w_n - w_n = \frac{E}{N} \geq \frac{\delta}{1+\theta} E$ for all n. Thus, starting from an outcome without evasion it is possible to reduce w_n by a small amount dw 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 $w_n, w_n \geq 0$, participation constraints $y_n = w_n - \tau w_n \geq y_n$, 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 $w_n - w_n \geq \frac{\delta}{1+\theta} E$ and $w_n \geq 0, w_n \geq 0$, $w_n - \tau w_n \geq y_n$ 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 $w_1 = \min_n w_n \geq y_n > 0$. Therefore, to maximize E, there is full evasion for the lowest-paid employee ($w_1 = 0$) and total evasion is taken to the point where (5) is binding for this employee, $E = \frac{1+\theta}{\delta} w_1 \geq N w_1$. All other employees evade by at least as much as the lowest-paid employee in absolute terms, $w_n - w_n \geq w_1$ 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 3. First, if $\delta = 0$, i.e., if the government o ers no reward for whistleblowing, then all rms will evade taxes (as in the standard model). Second, as soon as some reward $\delta > 0$ is o ered, then tax evasion is no longer sustainable for large rms. Therefore, the whistleblowing model also shows that low-powered nes and audit rates are enough to sustain truthful reporting in large rms. The key idea is the same as in the random shock model, namely that the inability to commit to not whistleblowing makes enforcement feasible.

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.

3.2 Firm Choice of Books vs. No Books

We now endogenize the decision to use books. For simplicity, we consider the random-shock model with parameter ε.

3.2.1 Reduced-Form Model

Various studies have shown that the informal sector is less e cient than the formal sector (see, e.g., La Porta and Shleifer, 2009, for a recent analysis), presumably because books help improve e ciency. We capture such e ects as follows. Let us assume that a rm with books has productivity $w = (w_1, ..., w_N)$, while a rm with no books has productivity $(1 - \alpha(N))w$ where $\alpha(N)$ represents the productivity loss of not using books. We assume that $\alpha(N)$ is increasing in N as books are more valuable when the business has more employees and is more complex.

There are three possible regimes associated with surplus levels Y_1, Y_2, Y_3. Denoting by $W = \sum_n w_n$, we have:

(1) No books and no taxes paid: $Y_1 = (1 - \alpha(N))W$

(2) Books and no income reported: $Y_2 = \{1 - [1 - (1 - \varepsilon)^N]\tau(1 + \theta)\}W$

(3) Books and incomes fully reported: $Y_3 = (1 - \tau)W$

As showed in Section 3.1.2, for a rm on the books, there is a threshold N such that $1 - \tau = 1 - [1 - (1 - \varepsilon)^{\bar{N}}]\tau(1 + \theta)$. For $N > \bar{N}$, the rm does not evade at all, whereas for $N < \bar{N}$, the rm fully evades and is audited with probability $1 - (1 - \varepsilon)^N$. The choice of books adds a layer of choice. If $Y_1 = \max(Y_1, Y_2, Y_3)$, then the rm does not use books and evades. If $Y_3 = \max(Y_1, Y_2, Y_3)$, then the rm uses books and reports truthfully. If $Y_2 = \max(Y_1, Y_2, Y_3)$, then the rm uses books but does not report any income.

It is reasonable to assume that, for very small N, $\alpha(N) < 0$, i.e., the cost of creating and maintaining books is higher than the productivity gain from using books. Conceivably, for very large N, $\alpha(N)$ converges to 1, i.e., the cost of not using books becomes prohibitive for large

businesses as it is impossible to carry out complex business operations without books. In this case, we have

Proposition 4 *Assume that the cost of not using books $\alpha(N)$ is increasing in N and such that $\alpha(N) < 0$ for small N and $\alpha(N) \to 1$ for large N. Then,*

(a) For small N, the firm does not use books and evades all tax

(b) For large N, the firm uses books and does not evade at all

(c) Depending on parameters, there may be an intermediate regime where the firm uses books and does not report any income

Proof: (a) follows from the assumption $\alpha(N) < 0$ for small N. (b) follows from the assumption $\alpha(N) \to 1$ for large N and Proposition 1. QED.

3.2.2 Micro-founding the Value of Books

Kremer (1993) sets out an O-ring theory of production. He considers a firm where N workers perform a series of N different tasks. If all tasks are performed perfectly, output per worker is B and total output is NB. But the product loses value at each stage, because workers perform the tasks less than perfectly. Mistakes in carrying out the task are multiplicative, and if one task fails, the product loses all its value. Worker n's skill at a task is denoted by q_n and is defined as the percentage of maximum value that the product retains when this worker performs the task. The production function is then given by $(q_1 \cdot ... \cdot q_N) \cdot N \cdot B$. Now, if ν is the fraction lost at each task due to not having books, a firm that is off-the-books retains a fraction $(1-\nu)^N$ of the on-the-books production value, i.e. $(1-\nu)^N \equiv 1 - \alpha(N)$ in the reduced form. This implies that the productivity loss from not having books increases in N. This result is driven by the presence of technical complementarity across production tasks and therefore generalizes beyond the specific O-ring production function.

Finally, by adding a fixed administrative cost c of creating books per unit of output, we would have $1 - \alpha(N) = (1-\nu)^N/(1-c)$. This specification satisfies the assumption stated in Proposition 4.

3.3 Additional Remarks on the Role of Books and Joint Information

Our theory posits that the success of third-party reporting derives from the presence of verifiable book evidence that is commonly known among a sufficiently large number of employees. It is

useful to contrast our theory with various situations where books are not necessary or present. First, consider the case of tips for restaurant workers. Tips are often additional o -the-books payments that take place directly between the restaurant clients and the restaurant employees. As a result, tips are typically not recorded on the books of the restaurant and they do not need to be: the restaurant owner knows that employees can supplement their on-the-books earnings with tips and can adjust on-the-book wages accordingly. Because tips are not on the book, even if a disgruntled employee denounces the restaurant, the books cannot provide a smoking-gun for cheating. Thus, tips are a way to escape third-party reporting, but they succeed only at the fringe in this very particular case where employees are directly in contact with clients to get a side payment.[14]

Second, consider the case of the corrupt o cial taking bribes from individuals or businesses to provide government services or to escape government regulation. This example may have the size factor as a corrupt o cial may receive bribes from many parties, but it does not have the hard book evidence as those bribes do not need to be recorded carefully in a \corruption accounting book". As a result, individuals or businesses denouncing the corrupt o cial might trigger an investigation, but the investigation is much less likely to succeed if there is no hard evidence beyond the denouncing claim.[15]

Third, a number of countries have enacted freedom of information legislation that requires the government to publicly disclose tax payer records, in some cases including exact tax return information. Some countries (e.g., Japan) limit such public records to top tax lers, while other countries (e.g., Norway and Sweden) provide public information for all tax payers. While such public disclosures are typically motivated by freedom of information rather than tax enforcement considerations, they may have important deterrence e ects on tax evasion as neighbors and acquaintances might ag serious under-reporting to authorities. Indeed, the presence of such public records extend joint information about reported income beyond the boundaries of the rm and increase the likelihood of whistleblowing. However, this type of enforcement is not as powerful as third-party reporting within rms, because whistleblowers cannot point to hard book evidence which makes it harder to detect and prove the existence of tax evasion. In addition,

[14] A modern government can easily crack down on tips by forcing service firms such as restaurants to post prices including services as most European countries do.

[15] An alternative is to use covered-up agents to catch corruption but this is costly and involves the risk of side collusion.

public tax records may create signi cant costs from loss of privacy.[16]

Finally, our theory has focused on labor income taxation, although in practice labor income taxes coexist with corporate pro ts taxes. It is often argued that the corporate pro ts tax helps the enforcement of the labor income tax, because wages are deductible in the corporate tax base. However, this argument implicitly assumes that total value added (pro ts plus wages) is veri able to the government. If total value added is not observable, then a rm can jointly under-report wages and pro ts. Typically, total value added will be observable only if the rm uses accurate accounting books. This is why, in our abstract and stylized model where employees and employers play a tax evasion game with the government, we do not need to make a distinction between employees earning wages and the employer earning pro ts.[17]

In sum, those examples illustrate what makes third-party enforcement successful. It is the combination of accurate and veri able books that is common knowledge within the rm and a large number of employees, as any single employee can potentially denounce collusive tax cheating by revealing the true books to the government. With many employees knowing about tax evasion but no veri able books, it would be much harder for the government to e ectively demonstrate that tax evasion took place. On the other hand, with books but only a few knowledgeable employees, it might be possible for the knowledgeable employees to collude and present false books to the government.

4 Allingham-Sandmo Model with Third-Party Reporting

In this section, we embed our agency model into a standard Allingham-Sandmo (A-S) model of tax evasion.

4.1 Irrelevance of Third-Party Reporting

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 w_t and w_s to the government. Based on our previous analysis, let us assume that we are in a situation with a large rm using books, implying that third-party

[16] Accounting books have the great advantage of being verifiable, necessary for business purposes, and not revealed to the public outside the firm.

[17] In many cases, there is an issue of classifying workers as employees vs. independent contractors. In general, withholding rules and information requirements are less for the latter, making evasion opportunities greater and forcing tax administrations to limit the scope of contractors (Logue and Slemrod, 2008).

reporting is truthful and hence $w_t = w_t$. In this case, self-reported income w_s is the only choice variable of the individual as in the A-S model.

The probability of being audited and having its under-reported income discovered is given by p. In the basic A-S model this probability p is independent of the reports w_s and w_t. The taxpayer solves the following expected utility maximization problem:

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

If w_s is not constrained to be positive (losses are fully deductible), we can rede ne the problem in terms of total reported income $w \equiv w_t + w_s$. The taxpayer then maximizes with respect to w, and the problem is exactly identical to the standard A-S formulation:

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

As the ne applies only to under-reporting, it is never optimal to over-report income so that $w \leq w$ and hence $(w - w)^+ = w - w$ in all that follows. Therefore, the FOC for w is given by

$$\frac{u'(c_A)}{u'(c_N)} = \frac{1-p}{p\theta}, \tag{8}$$

where c_A and c_N denote consumption in the audited and non-audited states. This gives a solution w^*, which is identical to the A-S model. Hence, we can state:

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 level of tax evasion is independent on the fraction of income w that is subject to third-party reporting.*

This implies that third-party reporting does not work, because the taxpayer can always 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 e ectively 100% crowd-out of self-reported income by exogenously increased third-party reported income.[18]

[18]This irrelevance result is formally similar to the 100% crowd-out of private provision of public goods by government provision.

4.2 Extended AS-Model: Breaking Irrelevance

There are two assumptions underlying the strong result in the previous section: (i) in the tax policy, losses are allowed to count fully against positive income in the tax calculation, (ii) in the audit policy, the audit probability does not depend on the report w_s, and in particular it does not depend on whether w_s is positive or negative. In real-world tax systems, these assumptions are unlikely to be satis ed. Deductibility of losses is disallowed to some degree and losses typically face higher audit rates than positive income. The above model o ers a possible explanation of why this is so: it is because governments wish to protect the enforcement bene ts of third-party reporting.

Disallowing Losses:

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

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

with respect to w_s. Now we have to distinguish explicitly between positive self-reporting ($w_s > 0$) and non-positive self-reporting ($w_s \leq 0$). As losses are disallowed, individuals with non-positive self-reporting ($w_s \leq 0$) are indi erent on how much losses to report, so we assume without loss of generality that $w_s = 0$ in that case. Therefore $w_s^+ = w_s$. Positive self-reporting is optimal if the right-derivative of (9) is positive at $w_s = 0$, 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 (characterized by eq. 8) was associated with $w^* > w_t$ (i.e., $w_s > 0$). 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 e ect on w. On other hand, once third-party reporting starts digging into previously unreported income (increasing w_t when $w_t \geq w^*$), then $w_s = 0$ and the individual can no longer o set third party reporting. We can summarize those results as follows:

Proposition 6 *Suppose that losses are disallowed and let w^* denote reported income in the standard A-S model. Consider a small increase in third-party reported income dw_t keeping total income w constant.*

(a) If $w_t < w^*$, then $dw_s/dw_t = -1$: increasing the fraction of income subject to third-party reporting does not increase total reported income $w_t + w_s$ (100% crowd-out).

(b) If $w_t > w^*$, then $dw_s/dw_t = 0$: increasing the fraction of income subject to third-party reporting increases one for one total reported income $w_t + w_s$ (0% crowd-out).

It would be worthwhile to develop a theory of optimal taxation in the presence of third party reporting.[19] Although we leave this analysis for future research, we conjecture that such a theory could justify disallowing self-reported losses or justify the use of schedular taxes, as we find in actual tax systems as we discuss below.

Differential Audit Rates:

Alternatively, we may consider the case where losses are fully deductible, but subject to a higher audit rate. In particular, we could assume $p = p_L$ if $w_s \geq 0$ and $p = p_H$ if $w_s < 0$, where p_H is greater than p_L. This is a special case of a general function, $p = p(w_s)$, that allows the audit probability to depend on the taxpayer report everywhere, not just at the threshold between gains and losses. To simplify the formal analysis, we assume that $p(w_s)$ is twice differentiable, and make the natural assumptions that $p'(w_s) < 0$ and $p''(w_s) \geq 0$. 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 w_s) lowers the risk of being audited.[20] The assumption of a non-negative second-order derivative reflects that, if the report is already unsuspicious (high w_s) and the audit probability therefore low, increasing reported income will have a relatively small impact.[21]

[19] Such an exercise would be similar in spirit to the important paper by Gordon and Li (2005) who show that the standard lessons of optimal tax theory no longer hold when part of the economic activity is informal and hence non taxable.

[20] 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'.

[21] 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.

The taxpayer maximizes

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

The first-order condition for w_s is given by

$$p(w_s)\, u'(c_A)\, \theta\tau - p'(w_s)\, [u(c_N) - u(c_A)] = (1 - p(w_s))\, u'(c_N)\, \tau, \tag{12}$$

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). Third-party reporting w_t now has an effect: consider an increase $dw_t > 0$ keeping w constant (so that $dw_s = -dw_t < 0$). If the taxpayer counteracts an increase in w_t by lowering w_s, there will be an effect through $p(.)$ and $p'(.)$.

Proposition 7 *Suppose that the audit probability $p(w_s)$ satisfies $p'(w_s) < 0$, $p''(w_s) \geq 0$ at the equilibrium (12). 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.*

Proof: Let us rewrite the first-order condition (12) to

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

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

$$\frac{dR}{dw_s} = \frac{\partial R}{\partial w_s} + \frac{\partial R}{\partial w} < 0.$$

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

$$0 = \frac{\partial R}{\partial w_s}[dw - dw_t] + \frac{\partial R}{\partial w} dw,$$

which implies

$$\frac{dw}{dw_s} = \frac{\partial R/\partial w_s}{\partial R/\partial w + \partial R/\partial w_s}.$$

The denominator is negative from the second-order condition. Hence $dw/dw_s > 0$ if and only if $\partial R/\partial w_s < 0$. As c_A, c_N do not depend on w_s, we have indeed

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

where the inequality follows from $p'(w_s) < 0$, $p''(w_s) \geq 0$, and $c_N \geq c_A$. QED.

Two points are worth noting. First, the condition $p''(w_s) \geq 0$ is sufficient (but not necessary) to ensure incomplete crowding out.[22] Second, the key reason why there is less than 100% crowd-out is because the government uses third party reports to determine its audit policy: p is a function of w_s and hence of $w_t = w - w_s$ as w is fixed. Indeed, the case $p(w)$ generates the same irrelevance result as in Proposition 5.

Finally and paralleling our discussion of Proposition 6, it would be worthwhile to develop a theory of optimal audit policy in the presence of third-party reporting, extending the model of Sanchez and Sobel (1993). We conjecture that the optimal audit policy would depend not only on total reported income but also on self-reported income.[23] We leave this analysis for future research.

4.3 Actual Income Tax Policies

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.[24] Losses in passive activities are also not allowed to count against other positive income items. Tax administrations also concentrate audits on self-reported income and especially self-reported income losses or deductions. 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 development of income taxation.[25] 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

[22] It is theoretically possible to obtain more than 100% crowding out if p is very concave.

[23] Recently, Bigio and Zilberman (2009) have extended the Sanchez and Sobel model to a situation where the government can also observe business size, which also generates a multi-dimensional auditing problem.

[24] 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.

[25] See Seligman (1911) for a comprehensive description of the history of income taxation.

bases. Initial comprehensive individual income taxes started with large exemption levels (so that only a minority of income 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 pro ts from large companies, which are all recorded on formal accounting books. As mentioned above, it would be worthwhile to develop a formal normative tax and audit theory framework which could generate as optimal policies the actual tax and audit policies observed over the course of economic development.

5 A Macroeconomic Theory of Tax Enforcement and Government Size

5.1 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 nanced 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 τ 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 has been normalized to one.

Government

We consider a benevolent government choosing public goods g and taxes τ 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.[26]

The government can convert one unit of c into one unit of g. Absent any enforcement

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

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 τ, 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 τ^* is characterized by:

$$1 = \frac{u_g(c,g)}{u_c(c,g)} = \frac{u_g(1-\tau^*, \tau^*)}{u_c(1-\tau^*, \tau^*)}. \tag{14}$$

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 $\hat{N}(A)$ denote the firm size that maximizes average productivity (minimizes average costs), i.e. $\hat{N}(A) = \arg\max_N x(N,A)$. This implies $x_N(\hat{N}(A), A) = 0$ and $x_{NN}(\hat{N}(A), A) < 0$, and 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 $x_N(\hat{N}, A) = 0$ and the assumption x_A/x increasing in N, which implies $x_{AN}(\hat{N}, A) = x_{NA}(\hat{N}, A) > 0$.

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 $\hat{N}(A)$.

5.2 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 3, 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, w) to maximize total surplus. Either they report truthfully ($w = w$) and workers pay taxes τw, or they report dishonestly ($w = 0$) 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.[27] Notice though, that in general equilibrium where free entry eliminates pure profits, the workers ultimately receive all the surplus from tax evasion.

We denote by y the net-of-tax income of each employee in his best outside option, where 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 $y/(1-\tau)$ if it complies with the tax law, and equal to y if it evades all taxes. Denoting by $1(w = w)$ the indicator variable that takes the value 1 under truthful reporting and value 0 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, the incentive to under-report wages to the government is that it 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 δ.

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, the income of each employee is given by $w = y = x(N, A)$. If an employee decides to whistleblow (given that nobody else does), he can obtain

[27] 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.

income $x(N,A) - \tau(1+\theta)x(N,A) + \delta\tau x(N,A)N$. The employee is therefore prevented from whistleblowing i $x(N,A) \geq x(N,A) - \tau(1+\theta)x(N,A) + \delta\tau x(N,A)N$, which is equivalent to $N \leq (1+\theta)/\delta$ as in Proposition 3. Hence, a rm that evades tax has to choose a rm size below $(1+\theta)/\delta$. Notice that the decision to whistleblow is independent of the level of public goods g, because whistleblowing within a single rm does not a ect the aggregate level of g.

Proposition 8 *We obtain the following cases:*

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

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

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

(b) If $x(\hat{N}(A),A)\cdot(1-\tau) \geq x((1+\theta)/\delta, A)$*, 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 pro ts are always greater under evasion when this can be sustained at the optimal rm size $\hat{N}(A)$. The proof of (2a) and (2b) follows from the observation that, once evasion is not sustainable under the optimal rm size $\hat{N}(A)$, an evading rm must reduce rm size to $(1+\theta)/\delta$. Under full evasion and $N = (1+\theta)/\delta$, the free-entry (zero-pro t) equilibrium is characterized by labor income $y = x(\frac{1+\theta}{\delta}, A)$. 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.

Note that Proposition 8 implies that taxation distorts rm size. The result is related to the empirical phenomenon of the \missing middle" discussed in the development literature (e.g., Tybout, 2000). In particular, Dharmapala, Slemrod and Wilson (2008) argue that the missing middle may be the outcome of optimal tax policies that exempt small rms from taxation in order to save on administrative costs. In our model, the missing middle does not arise because small rms are tax exempt *de jure*, but because small rms can sustain tax evasion and therefore become tax exempt *de facto*.

5.3 Macroeconomic Development and Optimal Government Policy

We now turn to the evolution of government size over the growth process. Let us denote by A_L the technology level such that $\hat{N}(A_L) = (1+\theta)/\delta$ and by A_H the technology level such that $x(\hat{N}(A_H), A_H) \cdot (1-\tau^*) = x((1+\theta)/\delta, A_H)$. Obviously, we have $0 < A_L \leq A_H$ and $A_L = A_H$ if $\tau^* = 0$. We can then state:

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 $x(\hat{N}(A), A) \cdot (1-\tau(A)) = x((1+\theta)/\delta, A)$. 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 proof is straightforward. The only non-obvious point is that $\tau(A)$ increases in A in the intermediate stage. Log-differentiating $1-\tau(A) = x((1+\theta)/\delta, A)/x(\hat{N}(A), A)$, we obtain

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

where we have used that $x_N(\hat{N}, A) = 0$. Because $\hat{N}(A) > (1+\theta)/\delta$ 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 accounting books of businesses, third-party reporting, etc.), while other taxes are the traditional taxes. Consistent with our theory, the graph shows clearly that virtually all of 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,[28] 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.

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

5.4 Endogenous Growth and the Possibility of Poverty Traps

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 basic education, health, and infrastructure, may be very important for economic growth. In the classical paper by Barro (1990), government inputs are complementary to private inputs in the production process, and sustained economic growth requires that a constant fraction of economic resources are devoted to government inputs. Without an expansion in government spending, growth comes to a halt because of decreasing returns to private capital. Hence, the government needs to raise government revenue in order to sustain economic growth. In Barro (1990), government spending is nanced by an income tax, and it is possible to solve for a xed tax rate, τ^*, that maximizes growth in private consumption, which is also the (second-best) optimal tax rate of a benevolent government. However, it is well-known that poor countries collect much less tax revenue as a fraction of GDP than richer countries and that personal income taxes account for only a minor part of the revenue in under-developed economies (e.g., Gordon and Li, 2005).[29] An explanation may be that it is impossible to enforce income tax collection in the early stages of development as described in Proposition 9. If our theory is combined with a Barro-type production technology with complementarity between private inputs and government inputs then an economy may get stuck in a poverty trap. In this case, economic growth comes to a halt because the net-return on private capital reaches the rate of time preference before the economy has entered into the intermediate stage of development in Proposition 9 where it is possible to raise tax revenue. On the other hand, if the economy enters into the intermediate stage, then the government will be able to raise revenue and sustain perpetual economic growth. In the intermediate stage, the tax rate is increasing and the growth rate of the economy is relatively high. At some point, the economy moves into the last stage after which the tax rate is τ^* and the growth rate of the economy is constant. Economists have proposed many theoretical mechanisms that may generate poverty traps (see Azariadis and Stachurski, 2005, for a survey). The public nance theory described above should be seen as complementary to those alternative theories.

[29] Poor countries rely to a larger extent on tariffs, seignorage and business taxation of large firms, which according to standard optimal tax theories are less effective sources of revenue.

6 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. 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.[30] 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 simple macro-economic growth model where the size and complexity of firms grows with exogenous technological progress. Our simple model can capture the stylized fact that the level of 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

[30] 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.

factors | size and complexity, links to the nancial sector, network of transactions | explains best the low levels of tax evasion observed in advanced OECD countries.

Our theory also predicts that it is the inability to tax the informal sector that is the main reason why developing countries have small governments as a share of GDP.[31] Other explanations have been put forward: (1) corruption of the tax administration may make taxes hard to collect in both the formal and informal sectors, (2) demand for government services might be lower in lower income countries. In principle, 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 corruption or demand explanations imply that even in the formal sector, tax rates should be lower in developing countries than in OECD countries.

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

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 (2008) \Tax structure, size of government, and the extension of the voting franchise in Western Europe, 1860-1938", *International Tax and Public Finance*.

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

Alm, J. and M. Mckee (2004). \Tax compliance as a coordination game," *Journal of Economic Behavior and Organization* 54, 297-312.

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.

Avi-Yonah, Reuven S. (2000) \Globalization, Tax Competition and the Fiscal Crisis of the Welfare State" Harvard Law Review 113, May 2000.

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 Endogeneous 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. (1996), \Children of the Performing Arts, The Economic Dilemma: The Climbing Costs of Health Care and Education." *Journal of Cultural Economics* 20, 183-206.

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 rms' 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.

Becker Gary S. and Casey B. Mulligan (2003) \Deadweight Costs and the Size of Government" *Journal of Law and Economics* 46 (October 2003).

Besley, Timothy and Torsten Persson (2007), \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 (2008) \State Capacity, Con ict and Development", LSE Working Paper, August 2008.

Bigio, Saki and Eduardo Zilberman (2009) \Optimal Business Income Tax Enforcement", NYU Working Papper.

Bird, Richard M. (1983), "Income tax reform in developing countries; the administrative dimension", *Bulletin for International Fiscal Documentation* 37, 3-14.

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

Bird, Richard. (2002) \Why Tax Corporations?" *Bulletin for International Fiscal Documentation* 52, 194-203.

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

Bird, Richard M. and Oliver Oldman (1975). *Taxation in Developing Countries*, John Hopkins University Press, 3rd Edition (1st edition is 1964).

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.

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

Frey, Bruno (1997) \A Constitution for Knaves Crowds Out Civic Virtues." *Economic Journal* 107, 1043-53.

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

Gordon, Roger and Wei Li (2005) \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 tari s: 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. (1996) *Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992*, IRS Publication 1415 (Rev. 4-96), Washington, D.C., 1996.

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), January 1981, 261-274.

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 April 2009.

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

Kremer, Michael (1993) \The O-Ring Theory of Economic Development Michael" *Quarterly Journal of Economics* 108(3), 551-575.

La Porta, R., Shleifer, A. (2009) \The Uno cial Economy and Economic Development," *Brookings Papers on Economic Activity*, forthcoming.

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

Levy, Santiago (2007), \Can Social Programs Reduce Productivity and Growth?" Working Paper, Eight Global Development Conference: Global Development Network January 2007.

Lindert, Peter H. (1994). \The rise in social spending 1880-1930." *Explorations in Economic History* 31, 1{37.

Lindert, Peter H. (2004). Growing public: social spending and economic growth since the eighteenth century (Volumes 1 and 2). Cambridge: Cambridge University Press.

Musgrave, Richard (1969) *Fiscal Systems*

OECD (2004). *Tax Administration in OECD: Comparative Information Series,* Organisation for Economic Co-operation and Development,

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

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

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). \Tari s and collections costs," *Review of World Economics* 123, 545{ 549.

Sanchez, Isabel and Joel Sobel (1993) \Hierarchical Design and Enforcement Income Tax Policies," *Journal of Public Economics* 50(3), 345-369.

Seligman, Edwin Robert Anderson (1911) *The Income Tax: A Study of the History, Theory, and Practice of Income Taxation at Home and Abroad*, New York, The Macmillan company.

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 Sel shness" *National Tax Journal* December, 57(4), 877{99.

Slemrod, Joel (2006) \Taxation and Big Brother: Information, Personalisation and Privacy in 21st Century Tax Policy," *Fiscal Studies* 27 (1) , 1{15.

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

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

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 (reproduced in Bird and Oldman 1975).

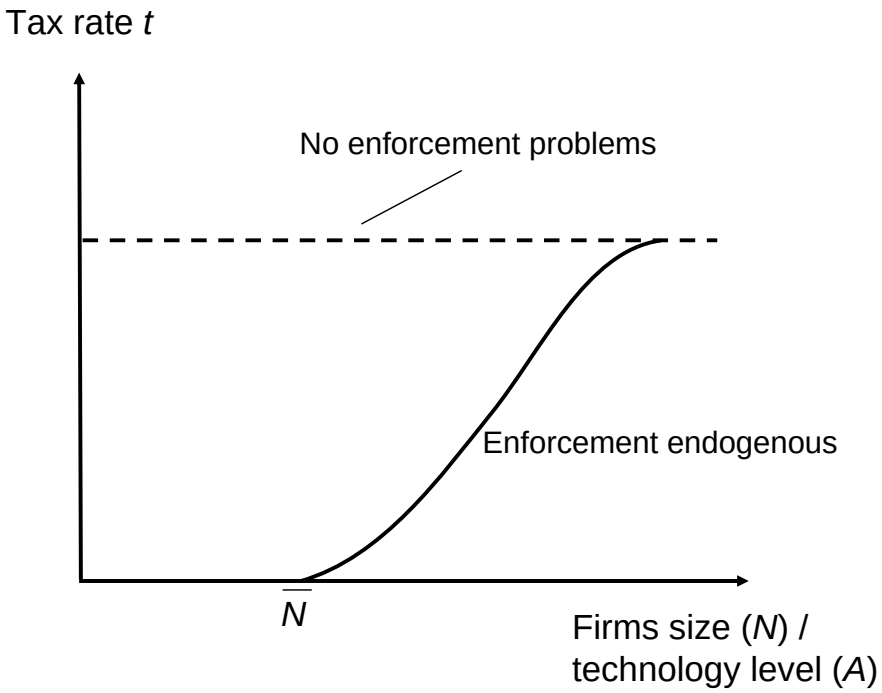
Tybout, James R. (2000). \Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature* 38(1), 11-44.

Webber, Carolyn, and Aaron B. Wildavsky (1986). *A history of taxation and expenditure in the Western World.* New York: Simon and Schuster.

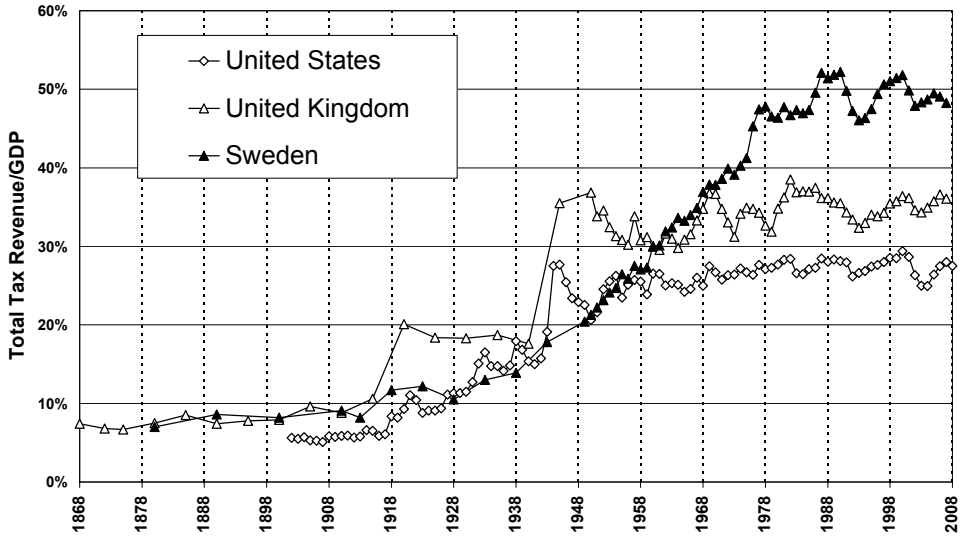
Yaniv, Gideon (1988) \Withholding and non-withheld tax evasion", *Journal of Public Economics* 35(2), 183-204.

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

Figure 1
Development in government size with and without tax enforcement problems



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



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

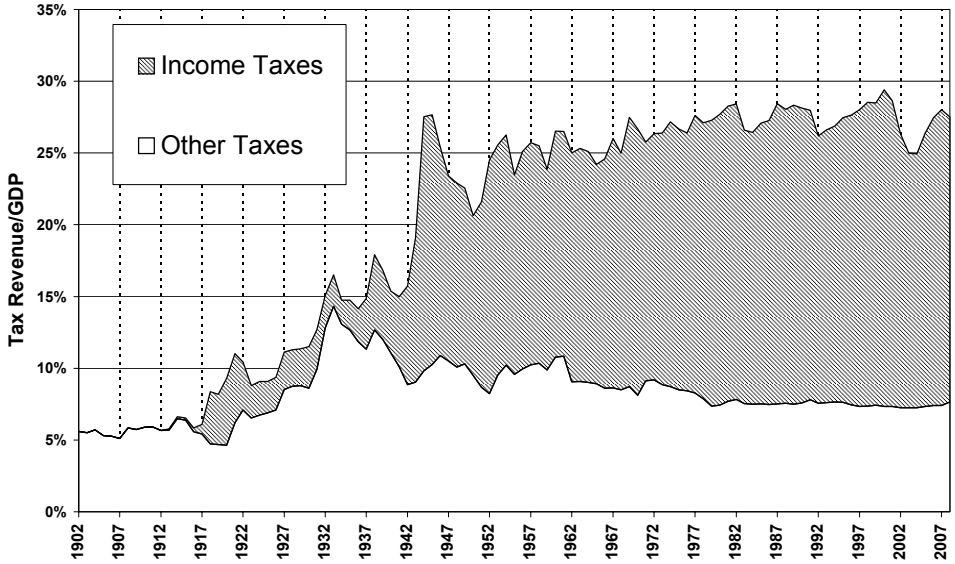


FIGURE 2
Evolution of Tax Revenue to GDP in Three Advanced Economies

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).