Economica

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

By Henrik Jacobsen Kleven†, Claus Thustrup Kreiner‡ and Emmanuel Saez††

†*London School of Economics*   ‡*University of Copenhagen*   ††*UC Berkeley*

We develop an agency model explaining why third-party information reporting by firms makes tax enforcement successful. While third-party reporting would be ineffective with frictionless collusion between firms and employees, collusive evasion is impossible to sustain in firms with many employees and accurate business records as any single employee may reveal evasion. We embed our agency model into a macro model where the number of employees grows with development, showing that the tax take evolves as an S-shape driven by changes in third-party information. We show that our model is consistent with a set of stylized facts on taxation and development.

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 income through taxation without destroying economic growth. In all advanced economies, most taxes are collected through third-party institutions such as employers, banks, investment funds and pension funds. These entities (which we call 'firms') generally have a large number of employees, clients or business partners. Therefore they need to use accurate and rigorous records to carry out their complex business activities. Firms report taxable income—such as compensation paid to employees and capital income paid to clients—directly to the government, and therefore act as a third party between households and the government. They also often withhold taxes on behalf of the government so that tax payments take place 'as-you-go'.[1]

It is widely known in the tax law literature (e.g. Surrey 1958; Lederman 2010) as well as among tax practitioners (e.g. 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 to a first approximation, tax enforcement is successful if and only if third-party reporting covers a large fraction of taxable income. For example, the most recent tax compliance study by the US Internal Revenue Service (2012) shows that the evasion rate for personal income is 56% when there is 'little or no' information reporting, while it is less than 5% when there is substantial information reporting. Kleven *et al.* (2011) obtain qualitatively similar results for Denmark.

However, it is not obvious why third-party reporting should work. Indeed, firms and employees have an incentive to collude to under-report income to the government and lower their tax bill. With frictionless collusion, third-party reporting cannot help tax enforcement, an important point made by Yaniv (1992). The goal of this paper is to develop a three-tiered agency model to provide a simple micro-foundation for the success of third-party reporting.

In our model, the government is the principal (top tier) 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 individuals. Although we focus on the case where individuals are employees of a firm, the model can easily be applied to situations where individuals are clients receiving capital income from a financial institution or shareholders receiving profits from a firm. When a firm is large and complex, using accurate business records—such as accounting books, details of purchases and sales, or payroll accounts listing individual wages and salaries—is extremely valuable for productivity. Such records are widely used within the firm, thus many employees know about them.

In principle and as pointed by Yaniv (1992), the firm and its employees could collude to report smaller incomes—salaries and profits—to the government than those actually earned. Under perfect information and commitment between the firm and individuals, there would be no reason for breaking the collusion. In practice, breakdowns can occur because of random shocks such as conflicts between employees and the employer, moral concerns of employees or an employee accidentally revealing the true business records to tax inspectors. Breakdowns can also occur as a result of rational whistleblowing if the government provides rewards to whistleblowers and firms cannot make employees commit not to whistleblow *ex ante*. In our model, we assume that each employee has the option of reporting cheating to the government by divulging the true business records to the government. When a firm has many employees, breakdowns of collusion will occur with a high probability. Critically, it is the combination of a large number of informed employees, and the existence of business records evidence, which makes third-party tax enforcement successful.[2] The assumption that all employees have information about business records is unrealistically strong. We show that our results largely carry over to the case where each employee is aware of only her own pay (reported and actual) and can denounce any discrepancy to the government as long as the government can find the business records in the case of an audit triggered by such a whistleblower. Again, it is the existence of business records that the government can eventually find that makes systematic tax cheating hard to sustain.

We embed our agency model into a simple macroeconomic growth model where the size and complexity of firms grow with exogenous technological progress. In this model, a representative individual has preferences over private and public goods. In the absence of enforcement problems, taxes are non-distortionary and should be set to finance public goods according to the classical Samuelson rule. We specify preferences such that the public good has an income elasticity equal to 1, implying that the first-best effective tax rate is constant along the path of economic growth. In the presence of enforcement constraints, there are three stages of development, and the tax-to-GDP ratio features an overall S-shape. In the earliest stage, firms are small and untaxable, therefore the government raises no tax revenue and supplies no public goods. In the middle stage, firms are large enough that they start becoming taxable provided that the tax rate is not too high. In this 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.[3]

We show that our macro model is consistent with a set of stylized facts on taxation and development. Gathering tax data for 14 advanced countries over a very long time period, we show that the historical evolution of the tax take is S-shaped in all countries and that the rise of taxation is entirely driven by third-party reported taxes such as personal income taxes and value-added taxes. While the exact timing of tax increases

varies somewhat across countries (depending for example on the exposure to wars), this stylized pattern holds everywhere. We also show that the well-known positive correlation between tax take and GDP per capita across countries at a point in time is driven entirely by modern third-party reported taxes, while there is no correlation with traditional self-reported taxes. Finally, we show that tax take and tax compliance are positively associated with firm size both across countries and across firms within a country. All of these findings are consistent with the predictions of our theory.

Our theory and evidence suggest that economic development is a necessary condition for the rise of large governments, with the transmission mechanism being the emergence of large and complex firms that can serve as third-party intermediaries and make it relatively easy to collect taxes from households. This implies that understanding the factors that shape tax capacity, including how to raise tax capacity in the current low-income countries, is closely related to understanding other aspects of the development process such as the change in firm structure.

Our paper relates to three literatures: (1) taxation and development, (2) theoretical tax compliance, and (3) growth of government. While we briefly discuss (1) and (2) here, we review the government growth literature in Section I in order to put our theory in context of the voluminous existing work on this central question.

Besley and Persson (2013, 2014) provide recent reviews of the literature on taxation and development. While third-party reporting covers most economic transactions in advanced economies, it is much more incomplete in developing countries. A number of recent empirical studies with compelling identification strategies have shown that tax enforcement in developing countries is affected by third-party information, providing tests for our theoretical model. Pomeranz (2015) analyses the role of third-party information for value-added tax enforcement in Chile. Randomized audit threats have much less impact on transactions that are subject to double reporting from both buyers and sellers, indicating that double reporting has a strong deterrent effect on tax evasion. Carillo *et al.* (2014) show, using a natural experiment in Ecuador, that there may be limits to the effectiveness of third-party information when taxpayers can make offsetting adjustments on less verifiable margins. Best *et al.* (2015), using evidence from Pakistan, show that turnover taxes can provide a useful alternative to corporate profit taxes, despite the production inefficiencies that they create, because sales are easier to observe than profits. Kumler *et al.* (2013) show that third-party enforcement of Mexican payroll taxes works better with larger firms. Naritomi (2015) shows that providing incentives for consumers to ask for value-added tax receipts and whistleblow non-compliant firms has large effects on reported value-added in Brazil. Cagé and Gadenne (2015) show that switches from tariffs to value-added taxes in developing countries may have led to reduced tax revenue, perhaps because these countries did not yet have the capacity to successfully enforce such modern, double-reported taxes.

The modern theoretical tax compliance literature grew out of Allingham and Sandmo (1972), which uses the Becker (1968) model of crime and focuses on a situation with no third-party reporting, that is, on the case where enforcement is never successful in practice and which covers a minor part of taxation in advanced economies. The Allingham–Sandmo model generates a key puzzle: why are compliance rates so high in developed countries given that audit rates and penalties for tax evasion are generally very low?[4] Relatively little attention has been paid to third-party reporting in the theoretical literature. A number of papers have shown that the classical results of optimal tax theory break down when tax enforcement is imperfect, and that this can help to explain observed tax structures in developing countries. Gordon and Li (2009) show that optimal

policies are closer to observed policies in a model where firms can evade taxes by avoiding the use of the financial sector. Best *et al.* (2015) show that in the absence of third-party information on profits, it may be socially optimal to use production inefficient tax instruments. Gadenne (2014) shows that food subsidies through ration shops along with commodity taxation can be welfare improving when governments have limited ability to observe household incomes due to the lack of third-party income reporting. Kopczuk and Slemrod (2006), Keen (2009) and De Paula and Scheinkman (2010) show that the equivalence between sales taxes and value-added taxes breaks down, because the value-added tax generates double reporting from buyers and sellers while the sales tax does not. Our paper focuses primarily on the within-firm information network rather than the across-firm information network and is therefore complementary to these studies.[5] Finally, a number of studies in the corporate income tax evasion literature have shown that the internal organization or the external activities of firms can affect their tax reporting decisions.[6]

The paper is organized as follows. Section I reviews the literature on government growth. Section II presents descriptive empirical evidence and develops a set of stylized facts about taxation and development. Section III sets out our micro model of third-party tax enforcement. Section IV embeds the micro model in a simple macroeconomic framework, which can account for the size and structure of taxation over the course of development. Section V concludes.

I. LITERATURE ON THE GROWTH OF GOVERNMENT

Our macro model contributes to a very long literature trying to explain the growth of government over the process of development. 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 1 (see, for example, Musgrave 1966). Second, Baumol's cost disease theory focuses on the supply side and posits that over the course of development, productivity in the private sector increases while productivity in the public sector stagnates, leading to a growth of government spending relative to GDP (Baumol and Bowen 1966; Baumol 1967). Third, Peacock and Wiseman (1961) propose a 'ratchet effect theory' whereby temporary shocks such as wars raise government expenditures, which do not fall back after the shock as social norms regarding the proper level of public goods and taxation are permanently affected by the temporary shock. Notice that the Wagner, Baumol and ratchet effect theories cannot explain the long period of stable government expenditures before the 20th century, a period with some economic growth and with many wars creating temporary spending shocks. Fourth, the Leviathan theory posits that governments are controlled by self-interested politician-bureaucrats, unchecked by electoral constraints (Brennan and Buchanan 1980), and hence maximize revenue under constitutional and fiscal constraints. Although proponents of the Leviathan theory have focused primarily on public choice and constitutional aspects, this theory is entirely consistent with the importance of tax enforcement constraints that we emphasize in this paper. Fifth, a large literature on political economy considers the role of voting, lobbying, corruption and political constitutions for the size of government. This literature has proposed that the democratization and increased political power of the poor have played an important role for the growth of government (Acemoglu and Robinson 2000). Moreover, substantial attention has been paid to the relationship between changes in income distribution and voters' demand for redistribution (Peltzman 1980; Lindert 2004).

In addition to these hypotheses, a number of studies have pointed out that there are fiscal capacity constraints to government growth (e.g. Kau and Rubin 1981; Bird 1992; Peltzman 1980; Riezman and Slemrod 1987; Kenny and Winer 2006; Aidt and Jensen 2009). Moreover, there is a vast literature on the role of underdevelopment in constraining tax structures both historically and in current developing countries.[7] Our theory proposes a micro-foundation that accounts for the changes in fiscal constraints over the course of development.

Recently, Besley and Persson (2009, 2010) have proposed an important extension of the ratchet effect theory that emphasizes the role of increasing fiscal capacity over the course of development. They develop a model where governments invest in fiscal capacity 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 for fiscal capacity investments in some countries such as the UK, we show in the next section that all advanced countries—including those that were not directly engaged in the major wars of the 20th century—have experienced the same stylized evolution of tax capacity, and that war involvement seems to be primarily related to the shorter-run timing of tax capacity investments. Furthermore, the question remains why recent (20th century) wars have led to large government expansions, whereas earlier wars typically did not. Our paper contributes to this question and is therefore complementary to the Besley–Persson theory.

II. Descriptive Empirical Evidence and Stylized Facts

To motivate our theoretical model, this section develops descriptive evidence on the cross-sectional and time series relationship between the level of economic development and the tax take. We emphasize the composition of the tax take into 'modern taxes' that rely on third-party information and 'traditional taxes' that tend to rely on self-reported information. Specifically, modern taxes are defined as personal and corporate income taxes, value-added taxes, payroll taxes and social security contributions, while traditional taxes are defined as all other taxes, which include property taxes, inheritance taxes, excise and sales taxes, custom duties, and so on.[8] Our data sources and exact definition of tax variables are described in an online appendix.

Figure 1 presents cross-country evidence. The data are from 2005 and include 29 countries within the OECD and 43 countries outside the OECD. Panel A depicts the well known positive correlation between GDP per capita and the total tax-to-GDP ratio: countries that have higher GDP per capita tend to have a much higher tax take. Panels B and C then split total taxes into modern and traditional taxes. Interestingly, while there is a clear positive correlation between GDP per capita and modern taxes to GDP, there is no correlation (or even a slightly negative correlation) between GDP per capita and traditional taxes. In other words, the relationship between taxes and development across countries is driven by a stark variation in tax *structure* across countries. We may state the following stylized fact.

Stylized Fact 1:

The positive relationship between tax take and economic development across countries is driven entirely by modern taxes that rely on third-party information and not at all by traditional taxes that rely on self-reported information.
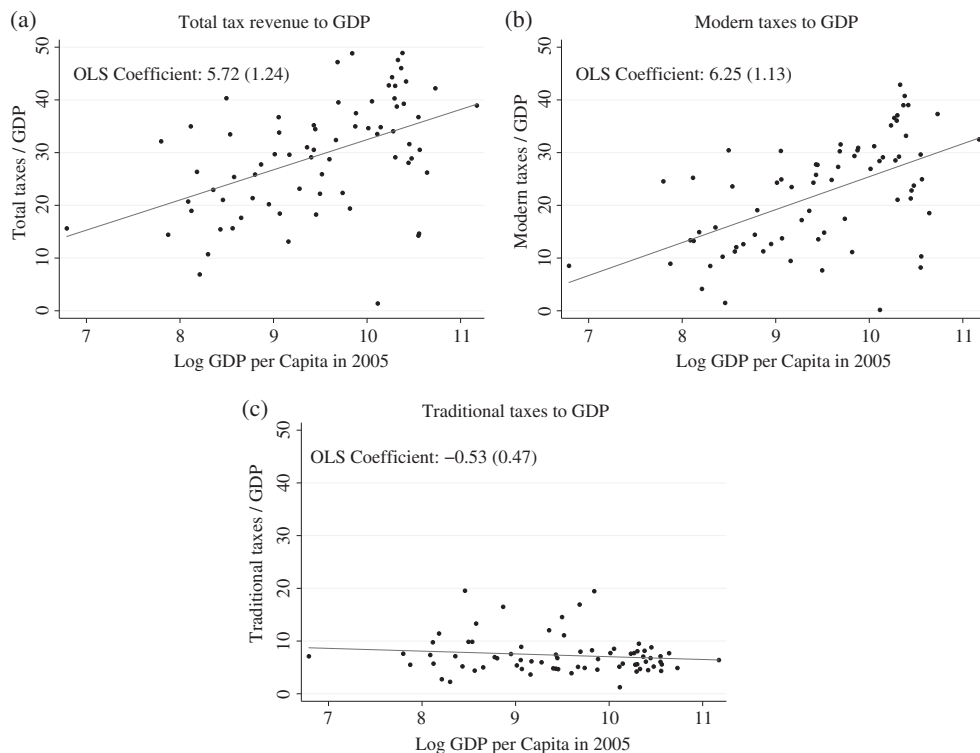
FIG. 1. Tax take and tax structure across countries.
Notes: IMF and OECD data from 72 countries, including 29 OECD countries as of 2005. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and value-added taxes. Traditional taxes include all the other taxes. See the online appendix for complete details.

Figure 2 and Figure A1 in the online appendix present time series evidence for 14 advanced economies for which data are available over the very long run, typically a century or more. We have constructed these series by combining the historical evidence from Flora (1983) along with the modern OECD (2008) *Revenue Statistics* series available since 1965 (all our constructed data are available online). Figure 2 focuses on four representative countries (France, Sweden, the UK and the USA), while the appendix figure shows the rest of the countries (Austria, Belgium, Denmark, Finland, Germany, Ireland, Italy, the Netherlands, Norway and Switzerland). Each panel plots, for a given country, the time series of the total tax-to-GDP ratio and decomposes it into traditional taxes (light shading) and modern taxes (dark shading). Three points are worth noting. First, all countries display an overall S-shape for the tax-to-GDP ratio. The tax take was small until about a century ago (typically less than 10% of GDP), increased sharply during the 20th century, and then stabilized from around the late 20th century (at 35–50% depending on the country). Second, the growth in tax take is driven entirely by growth in modern taxes, with no secular increase in traditional taxes (and typically a weak decline). Third, the exact timing of the tax increase differs across countries. For example, most of the increase takes place around the world wars in the UK. The USA also displays clear spikes around the world wars, although the tax ratio comes down to some extent after the wars. On the other hand, the increase in tax take is very smooth in France and Sweden (the latter being relatively unaffected by the wars due to its status as a
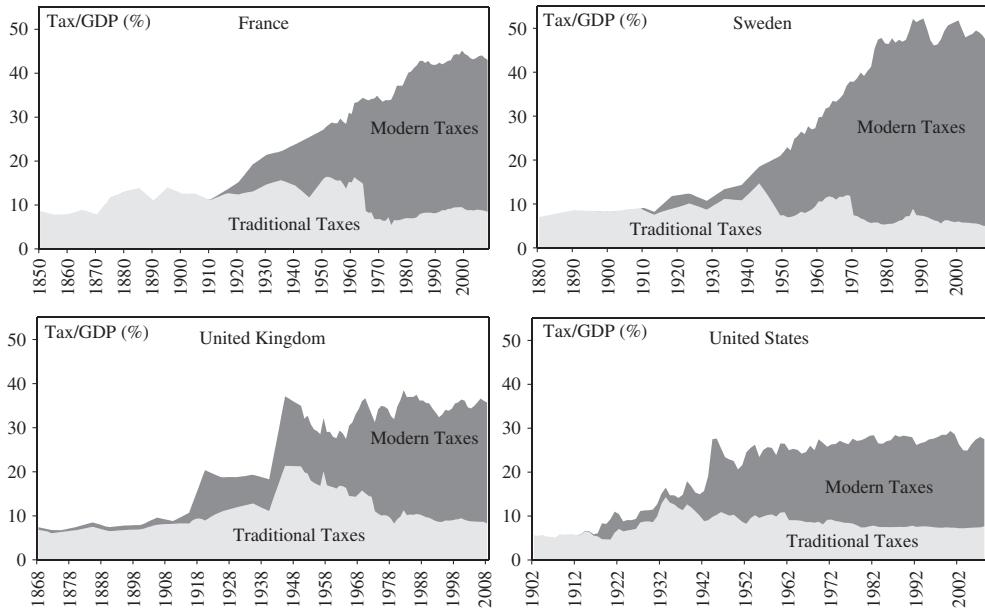
FIG. 2. Tax take and tax structure over time.

Notes: Data from Flora (1983) before 1965 and from OECD after 1965. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and value-added taxes. Traditional taxes include all the other taxes. See the online appendix for complete details.

neutral country) as well as in most of the other countries shown in the appendix.[9] In all countries and despite their different exposures to wars, the stylized pattern of government growth follows an S-shape driven by the expansion of third-party enforced taxes. We may then state our next stylized fact.

Stylized Fact 2:

The evolution of the tax take over the course of economic development within countries follows an S-shaped pattern, with the increase in tax take driven entirely by modern taxes while traditional taxes remain constant or weakly falling.

Figure 3 explores a potential mechanism that we highlight in our theoretical framework, namely the numbers of employees in firms. Panel A plots the tax-to-GDP ratio against the share of workforce employed in large firms (defined as firms with 10 or more employees) across countries based on data from the Global Entrepreneurship Monitor (GEM) survey conducted in more than 50 countries (see, for example, Poschke 2014). The graph shows that tax take and share of workforce in large firms are positively correlated across countries.[10] We may therefore state our final stylized fact.

Stylized Fact 3:

Tax take across countries is positively related to the share of employees working in large firms.

To examine the within-country effect of firm size on tax evasion, Panel B of Figure 3 shows estimates of tax evasion rates among Danish firms by number of employees. These
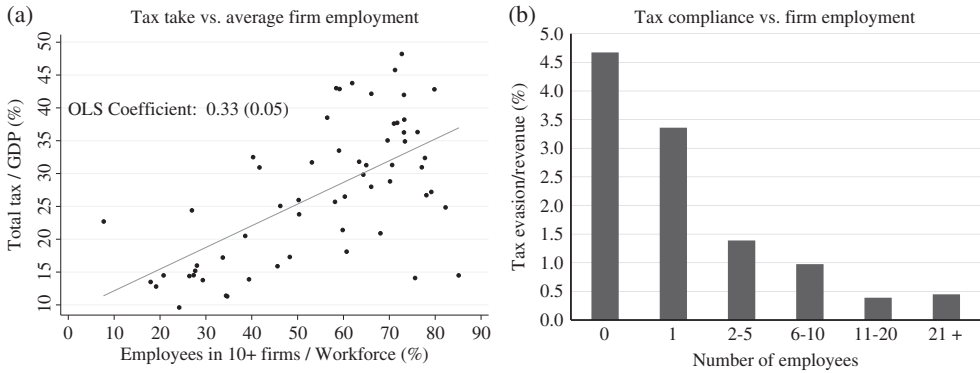
FIG. 3. Tax take and tax compliance vs. firm size.
Notes: Panel A shows total tax to GDP ratio in 2012 (source: Index of Economic Freedom, Heritage Foundation) by the share of the workforce working in firms with 10+ employees (source: Global Entrepreneurship Monitor survey and World Bank data) across 59 countries where information exists. Panel B shows tax evasion rates among Danish firms by number of employees, and is based on random audits carried out by the Danish Tax Agency in 2007 (source: data from the Danish Tax Agency SKAT). See the online appendix for complete details.

estimates are based on randomized tax audits of firms implemented as part of a large-scale field experiment in Denmark (see SKAT (2009) for the official report and Kleven *et al*. (2011) for an analysis of the individual tax component of this experiment). The evasion includes all forms of tax evasion by the firm, including corporate tax evasion on profits (e.g. under-reporting sales or over-reporting costs) but also tax evasion of labour costs (e.g. cash wages paid under the table with failure to withhold income and payroll taxes on such wages). The figure shows a stark negative relationship between tax evasion and the number of employees, suggesting that within a country, tax compliance by firms is related to firm size. Unfortunately, the data from SKAT (2009) do not break down tax evasion for corporate tax vs. tax evasion for labour income taxes so that panel B of Figure 3 cannot zoom in specifically on tax evasion for labour income taxes by third-parties.[11]

The next two sections develop a micro–macro model that is consistent with the stylized facts presented above.[12]

III. A MICRO THEORY OF THIRD-PARTY TAX ENFORCEMENT

Let us assume that N individuals are working in a firm and receive pre-tax wages $w = (w_1, \ldots, w_N)$. The pre-tax profits of the firm are denoted by Π. Hence the total value-added created by the firm is equal to $V = W + \Pi$, where $W = \sum_n w n$ are aggregate wages in the firm. Value-added is also equal to total sales S minus purchases P. Let us assume that the government imposes a flat tax at rate τ on both wages and profits. If S and P are observable to the government, then value-added $V = W + \Pi = S - P$ is also observable. As a result, under-reporting wages is useless to the firm because this would automatically increase its tax on profits.[13] However, if S and P are not observable to the government, then the firm can possibly under-report wages W without having to over-report profits Π.[14]

In practice, S, P and W (and hence Π) would be observable to the government if the firm truthfully records this information in its business records (such as accounting books

and payroll lists) and the government has access to these business records. Some firms may be able to carry out their business without recording this information formally. For example, a small family business might carry out all or part of its purchases and sales with cash and never record this information. On the other hand, maintaining accurate business records is clearly helpful to firm productivity: the business can measure its profits accurately, keep track of wages paid out, plan production activities, and obtain access to financial sector services, formal insurance, etc. Realistically, the productivity gain of keeping business records is larger when the firm is larger and more complex, and for modern firms the cost of being off-the-books becomes prohibitive. We therefore assume that the firm maintains accurate business records, which creates potentially detectable information within the firm.[15] However, even though business records exist, the firm may still be able to hide those records from the government in order to evade taxes. For example, the firm may maintain a double set of books—true books for business purposes and edited books for tax purposes. In this section, we present a simple agency theory showing how the government can truthfully extract the true business record information using third-party reporting.[16]

Because we assume that the tax rate τ on profits and wages is the same, there are no incentives for profits and wage shifting, and therefore wages and profits can be treated symmetrically. Hence without loss of generality, we can model the owner of the profits as one additional wage earner, which simply amounts to ignoring profits (setting $\Pi \equiv 0$) in the analysis.[17]

Basic setup

We assume that the government sets in place third-party reporting for tax purposes whereby each employee is required to report her earnings to the government and the firm is also required to report such individual earnings directly to the government.[18] Therefore employees and employers have to agree on a wage report to the government, as any discrepancy in the employer and employee reports would generate a tax audit.[19]

We can therefore assume that the firm and employees agree on reports to the government given by $\overline{w} = (\overline{w}_1, \ldots, \overline{w}_N)$, and this determines tax payments to the government unless any tax cheating is detected. We consider a situation where both real and reported wages $(w, \overline{w})$ are determined cooperatively by the N employees of the firm. Because this is a tax collusion game, a cooperative game seems to be the most natural one.[20] As a 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 $\overline{y} = (\overline{y}_1, \ldots, \overline{y}_N)$ the disposable income levels (net of taxes) associated with those outside options.[21] In the general equilibrium macro-model presented in Section IV, we fully endogenize outside options and firm size N.

The presence of business records creates common knowledge within the firm. We capture such common knowledge by assuming that $(w, \overline{w})$ is known to everyone within the firm. In practice, although records may not be known to literally everyone within the firm, they are widely used in the firm and will be known by a number of employees. We explore also the alternative polar case where only employees for whom $w_n \neq \overline{w}_n$ are aware of tax evasion and can denounce tax cheating within the firm. This situation of

private knowledge of tax evasion might be more realistic in the case of *external* parties such as business or individual clients, shareholders or debt holders, a point to which we return later.

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

When evasion is detected, we assume that the government charges the evaded tax plus a fine. As in all tax enforcement studies, we assume that there is an exogenous upper bound θ on the level of fines relative to tax evaded.[22] In that case, it is straightforward to show that it is always best for the government to impose the maximum possible fine in all circumstances. Therefore without loss of generality, we assume that the penalty is equal to $\theta\%$ of the evaded tax to each person caught evading. In addition, the government may offer a reward to whistleblowers equal to a share δ of total uncovered tax evasion. For simplicity, we assume that all workers are risk-neutral.

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

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

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

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

introducing a *trembling hand*, while the whistleblower model shows that the evasion equilibrium is not robust to relaxing the *perfect commitment* assumption.

Random shock model

We incorporate the possibility that an employee may deviate and reveal internal business records either by mistake, because he is disgruntled, or because of moral concerns.[23] Let ε be the probability of any given employee revealing true information through such random shocks. We assume for simplicity that those shocks are independent and identically distributed across employees. With N employees, nobody will denounce tax cheating with probability $(1-\varepsilon)^N$. The probability that somebody in the firm reveals true information (and hence triggers an audit) is therefore given by $1-(1-\varepsilon)^N$. This probability is increasing in N, and tends to 1 as N tends to infinity as a random shock is bound to happen when the number of employees is very large.

The expected pay-off of each employee equals

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

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

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

When $\overline{w}_n > w_n$, we have

$$\frac{\partial Y}{\partial \overline{w}_n} = -\tau,$$

so it never pays to over-report wages.[24]

Proposition 2. In the random shock model, any cooperative solution satisfies the following.

(a) If $(1-\varepsilon)^N \leq \theta/(1+\theta)$, then there is no tax evasion at all: $\overline{w} = w$.
(b) If $(1-\varepsilon)^N > \theta/(1+\theta)$, then there is complete tax evasion: $\overline{w} = 0$.
(c) For any $\theta > 0$ and $\varepsilon > 0$, there is $\overline{N}$ such that firms do not evade when $N \geq \overline{N}$.

Proof. The proof of (a) and (b) is immediate as $\partial Y/\partial \overline{w}_n \geq 0$ if and only if $(1 + \theta)(1 - (1 - \varepsilon)^N) \geq 1$ if and only if $\theta/(1+\theta) \geq (1-\varepsilon)^N$. For (b), where $\partial Y/\partial \overline{w}_n < 0$, the

solution is determined by the non-negativity constraint $\overline{w}_n \geq 0$ for all n. For (c), $\overline{N}$ is defined by $\theta/(1+\theta) = (1-\varepsilon)^{\overline{N}}$, that is, $\overline{N} = \log(\theta/(1+\theta))/\log(1-\varepsilon)$.

Four points are worth noting about Proposition 2. First, when $\varepsilon = 0$, we are back to the standard collusive case where firm size does not help and there is always tax evasion. Second, when $\varepsilon > 0$ and even for moderate fines $\theta > 0$, it will always be the case that large firms choose not to evade, destroying the evasion equilibrium from Proposition 1. This is consistent with Stylized Fact 3 from Section II, showing that tax enforcement works better in large firms. Our model can therefore explain why low fines and low audit rates can lead to successful enforcement in practice. This resolves the key puzzle of the Allingham–Sandmo model, which predicts extremely high evasion rates when audit rates and fines are low (given reasonable risk-aversion parameters). Third, our qualitative results are robust to introducing risk aversion, which would make tax enforcement easier. Fourth, the results in Proposition 2 do not depend on the specific division of revenue W across workers. The equilibrium division will depend on the outside opportunities $\overline{y}$ and other factors not explicitly specified that determine the bargaining power of the individuals.

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

Proposition 3. In the random shock model with only private information on incomes, we have the following.

(a) The optimal evasion strategy for the firm is to report zero income for the N_c highest-paid employees, where N_c is an integer below $\overline{N}$ defined as $[1 - (1-\varepsilon)^{\overline{N}}](1+\theta) = 1$.

(b) Assuming a fixed distribution of wage incomes, the fraction of income evaded tends to zero as N gets large.

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

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

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

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

Therefore evasion is profitable only if $N_c \leq \overline{N}$ is defined as $N_c = [1 - (1-\varepsilon)^{\overline{N}}](1+\theta) = 1$. An equilibrium with $N_c \leq \overline{N}$ evaders Pareto-dominates an equilibrium with truthful reporting, because the payoff from the N_c evaders is higher due to under-reporting, while the payoff from everybody else is unaffected. Moreover, when an employee evades, the surplus is maximized by full evasion: $\overline{w}_n = 0$. Because the extra surplus created by full evasion is proportional to w_n, surplus is maximized by having the highest-paid employees evade. Given $N_c \leq \overline{N}$, the optimal number of evaders reflects a trade-off between the extra surplus from the N_cth evader and the higher probability of being caught for all other evaders. It is optimal to evade for at least one employee (the highest paid) if and only if $\varepsilon(1+\theta) \leq 1$, that is, if and only if $\overline{N} \geq 1$.

(b) Because $\overline{N}$ is fixed, as N goes to infinity we have that $N_c/N \leq \overline{N}/N$ goes to zero—a vanishing fraction of employees will be able to evade. If the wage distribution is fixed, then the share of total compensation going to a vanishing fraction of employees also converges to zero.

Two points are worth noting about Proposition 3. First, our result of successful enforcement for large firms remains valid in the case of only private information, which is the least favourable to tax enforcement. Second, this case may capture some of the real-world tax evasion practices of large firms. Most of the corporate income tax evasion does not take place as collusion to under-report the wages of ordinary employees, but rather 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 Proposition 3 (see, for example, Slemrod 2004). Firms that plan to evade taxes therefore have an incentive to limit the flow of information within the firm.

Rational whistleblower model

We now consider the case where the government offers a whistleblower reward and we assume that each individual may voluntarily and rationally denounce their employer. Hence we relax the critical assumption of *ex ante* commitment from Proposition 1. In practice, firms do not have the power to enforce non-whistleblowing commitments.[25] We assume that the whistleblower reward is equal to a fraction δ of total uncovered revenue shared among all whistleblowers.[26] In theory, the government can choose δ, and a higher δ helps compliance. Therefore the government would want to make δ as high as possible. The same phenomenon arises in the classical tax evasion of Allingham and Sandmo (1972), where increasing the fine rate for evading taxes always helps compliance and hence should be set as high as possible. In practice, there are fairness limits on how high rewards or punishments can be. Therefore in our analysis, we consider $\delta > 0$ as exogenous and we show that our theory carries over even if δ is small, so our results do not rely on the use of high-powered but unrealistic incentives.

Several OECD countries use such whistleblower rewards to induce insiders to denounce large-scale tax evasion within firms. For example, in the USA, 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). More recently, in 2012 Bradley Birkenfeld, a former UBS AG banker, secured a whistleblower award of $104 million, the largest individual federal payout in US history, after telling the Internal Revenue Service how the bank helped thousands of Americans to evade taxes (Schoenberg and Voreacos 2012). Relatedly, Japan allows laid-off workers to claim unemployment benefits even if their employer did not pay social security contributions (OECD 2004). Such claims help the government to discover businesses evading social security taxes.[27] 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 to uncover creates a psychological reward of δ dollars.[28]

Given payments $w = (w_1,\ldots,w_N)$ and reports $\overline{w} = (\overline{w}_1,\ldots,\overline{w}_N)$, the payoff for employee n if he does not whistleblow is given by

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

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

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

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

From equations (2) and (3), the total surplus in the firm can be written as

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

A cooperative solution $(w,\overline{w})$ maximizes surplus Y subject to $\sum_{n'} w_{n'} = W$, non-negativity constraints $w_n, \overline{w}_n \geq 0$ for all n, and participation constraints $y_n \geq \overline{y}_n$ for all n. Notice that $(1-\delta)(1+\theta) \geq 1$ if and only if $\delta \leq \theta/(1+\theta)$ is required to avoid a situation where employees always evade and then collectively whistleblow in order to recoup larger rewards than the fines that 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 interests to whistleblow *ex post*. Therefore given that co-workers do not whistleblow, utility for employee n must be higher under no whistleblowing (equation (2) with $a = 0$) than under whistleblowing (equation (3) with $N_w = 1$), implying that for all n,

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

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

Proposition 4. In the whistleblower model, any cooperative solution satisfies the following.

(a) If $N > 1/\delta$, then there can be no tax evasion at all: $\overline{w} = w$. Hence large firms do not evade taxes even if $\delta > 0$ is very small.
(b) If $N \leq 1/\delta$, then some evasion is sustainable, and an outcome without evasion is Pareto-dominated by a sustainable evasion equilibrium. In the evasion equilibrium, the lowest-paid employee always reports zero wages (full evasion). All other employees may report positive wages (less than full evasion), but evade by at least as much as the lowest-paid employee in absolute terms. If wages $w_1, \ldots, w_N$ are equal, then all employees report zero wages.

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

For (b), if some evasion is sustained ($E > 0$), then we must have $w_n - \overline{w}_n \geq \delta E$ for all n. Because $\delta \leq 1/N$ in this case, it is feasible to satisfy this condition, for example by having equal evasion across all employees: $w_n - \overline{w}_n = E/N \geq \delta E$ for all n. Thus starting from an outcome without evasion it is possible to reduce $\overline{w}_n$ by a small amount $d\overline{w}$ for all n and thereby generate a sustainable Pareto improvement. The evasion equilibrium is characterized by the maximization of total surplus Y at $a = 0$ subject to $\sum_{n'} w_{n'} = W$, non-negativity $w_n, \overline{w}_n \geq 0$, participation constraints $y_n = w_n - \tau \overline{w}_n \geq \overline{y}_n$, 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 $w_n - \overline{w}_n \geq \delta E$ and $w_n \geq 0$, $\overline{w}_n \geq 0$, $w_n - \tau \overline{w}_n \geq \overline{y}_n$ 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 evade by at most $w_1 = \min_n w_n \geq \overline{y}_n > 0$. Therefore to maximize E, there is full evasion for the lowest-paid employee ($\overline{w}_1 = 0$), and total evasion is taken to the point where (5) is binding for this employee, $E = w_1 / \delta \geq N w_1$. All other employees evade by at least as much as the lowest-paid employee in absolute terms, $w_n - \overline{w}_n \geq w_1$ 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.

Three points are worth noting about Proposition 4. First, if $\delta = 0$—that is, if the government offers no reward for whistleblowing—then all firms will evade taxes as in Proposition 1. Second, as soon as some reward $\delta > 0$ is offered, tax evasion is no longer sustainable for large firms. Therefore the whistleblowing model also shows that low-powered fines and audit rates are enough to sustain truthful reporting in large firms (consistent with Stylized Fact 3 from Section II). This shows that the collusion equilibrium of Proposition 1 is not robust to relaxing the assumption of perfect commitment. Third, in this model, equality in the distribution of true wages $w_1, \ldots, w_N$ 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 that it delivers.

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

Mechanism design

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

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

Second, if there is more than one individual ($N \geq 2$), then in principle the government could design a non-conventional whistleblowing mechanism that induces truthful reporting. This mechanism is as follows: if the government receives information from N_w whistleblowers, then it will randomly select one whistleblower n^*, forgive n^* his evaded tax and corresponding fine and offer n^* a small fraction of the tax evaded by the other individuals.[29] This mechanism would induce any individual to denounce tax cheating and make tax collusion impossible to sustain as long as $N \geq 2$. This strong implementation result is consistent with the mechanism design literature, which has shown that first-best is often implementable in common information environments using sufficiently sophisticated mechanisms (Moore 1992).

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

External business records and the scope of the firm

Our theory posits that the success of third-party reporting derives from the presence of verifiable *internal* business records that is commonly known among a sufficiently large

number of employees. It is useful to contrast our theory with situations where such records are not present, or when *externally* recorded transactions allow outside business partners to denounce the firm.

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

The presence of publicly disclosed accounting books certainly imposes constraints on how much firms can evade, as accounting books and corporate tax returns must be consistent. Theoretically, the firm could collude with shareholders and banks to publicly disclose fake accounting books while secretly showing the true books to prospective shareholders and lenders. Exactly as in our model, such collusion would be very difficult to maintain with a large number of players. Therefore firms that want to raise equity or debt need to maintain accurate business records and cannot easily escape taxation.[30] If taxes on earnings are not linear, then it is still possible to manipulate the distribution of reported earnings while truthfully reporting total earnings. This type of evasion could be analysed along the lines that we have proposed.

Value-added taxes (VAT) require firms to keep accounts of all purchases and sales, and pay taxes on sales net of purchases. Therefore each firm has an incentive to under-report sales and over-report purchases, hence creating opposite incentives across businesses engaged in arm-length transactions (De Paula and Scheinkman (2010) present a simple theoretical model of VAT enforcement with a formal/informal decision from firms). Starting from a no-evasion equilibrium, only businesses selling directly to households for final consumption can unilaterally evade by under-reporting sales. Even in that case, evasion is partial as businesses cannot consistently report negative value-added without raising suspicions. Exactly as in our model, we would expect small retailers to be able to partly evade the VAT, while large retailing chains need to maintain formal business records, making evasion much harder. Businesses further up in the VAT chain need to collude with businesses further down the chain in order to evade VAT. Therefore as long as there is a large business further down the chain, VAT evasion is not feasible even for small informal businesses (for an empirical analysis, see De Paula and Scheinkman 2010; Pomeranz 2015). However, if all businesses were small and informal, it would be impossible to implement a VAT as the tax would unravel from the bottom up. Therefore in the end, we believe that it is again the presence of a large business that uses business records and cannot successfully hide them that makes the VAT successful, exactly as in our basic model.[31]

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

IV. A Macro Theory of Tax Enforcement and Government Size

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

Macro model without enforcement problems

Households There is a continuum (of measure 1) of homogeneous individuals, who derive utility $u(c,g)$ from the consumption of a private good c and a tax financed public good g. We assume that $u(c,g)$ is homothetic, implying that the public good has an income elasticity equal to 1 (see below). We also assume that $u_c(c,0) > 0$, so that public goods are not essential for prosperity.

We assume that labour is inelastically supplied. We denote by w the pre-tax labour 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 is normalized to 1.

Government We consider a benevolent government choosing public goods g and tax rate 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.[32]

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

$$(6) \qquad 1 = \frac{u_g(c,g)}{u_c(c,g)} = \frac{u_g(1 - \tau^*, \tau^*)}{u_c(1 - \tau^*, \tau^*)}.$$

Importantly, the optimal effective 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 1. This implies that the ratio 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 labour 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 labour input, defined as $x_A(N,A)/x(N,A)$ being increasing in N.

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

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

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

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 $\widehat{N}(A)$.

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

Incorporating tax evasion into the model

We consider the whistleblower model of tax evasion that we view as a simple, reduced-form model of tax enforcement that could possibly stand in for other mechanisms that we discussed in earlier sections. In particular, the whistleblower model simplifies the presentation because it involves no uncertainty. From Proposition 4, either there is evasion that always goes undetected or there is no evasion at all. Furthermore, because all workers are identical in this model, when there is evasion, it is complete.

As before, we consider a cooperative game where the firm and its employees agree on true and reported wages $(w, \overline{w})$ to maximize total surplus. Either they report truthfully $(\overline{w} = w)$ and workers pay taxes τw, or they report dishonestly $(\overline{w} = 0)$ 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 in the micro model in Section III, we characterize not the entire set of cooperative equilibria (the core), but a specific equilibrium where the firm gets the surplus from evasion.[33] Notice, though, that in general equilibrium where free entry eliminates pure profits, the workers ultimately receive all the surplus from tax evasion.

Let $\overline{y}$ be the net-of-tax income of each employee in his best outside option, where $\overline{y}$ is determined by the equilibrium in the labour market and taken as given by the firm. The firm has to offer each employee a pretax compensation equal to $\overline{y}/(1-\tau)$ if it complies

with the tax law, and equal to $\overline{y}$ if it evades all taxes. Denoting by $1(\overline{w} = w)$ the indicator variable equal to 1 under truthful reporting and 0 under full evasion, profits can be written as

$$x(N, A) \cdot N - \frac{\overline{y}}{1 - \tau \cdot 1(\overline{w} = w)} \cdot N.$$

Hence for the firm, under-reporting wages to the government lowers the before-tax wage that 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 subsection that equilibrium firm size equals $\widehat{N}(A)$, the before-tax wage is given by $w = x(\widehat{N}(A), A)$, and the after-tax wage is given by $y = (1 - \tau) \cdot x(\widehat{N}(A), A)$. If the firm evades and nobody whistleblows, then each employee's income is $w = y = x(N,A)$. If an employee whistleblows (and nobody else does), then he obtains $x(N,A) - \tau(1 + \theta)x(N,A) + \delta\tau(1 + \theta)x(N,A)N$. Therefore the employee does not whistleblow if and only if $x(N,A) \geq x(N,A) - \tau(1+\theta)x(N,A) + \delta\tau(1 + \theta)x(N, A)N$, which is equivalent to $N \leq 1/\delta$ as in Proposition 4. Hence a firm that evades tax must choose a firm size below $1/\delta$.[34]

Proposition 5. We obtain the following cases.

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

   (a) if $x(\widehat{N}(A), A) \cdot (1 - \tau) < x(1/\delta, A)$, then the firm evades all taxes and chooses suboptimal firm size $1/\delta$;
   (b) if $x(\widehat{N}(A), A) \cdot (1 - \tau) \geq x(1/\delta, A)$, then the firm does not evade and chooses the optimal firm size $\widehat{N}(A)$.

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

   Note that Proposition 5 implies that taxation distorts firm size away from intermediate levels above $1/\delta$. The result is consistent with the empirical phenomenon of the 'missing middle' discussed in the development literature (e.g. Tybout 2000).[35] Dharmapala *et al.* (2011) argue that the missing middle may be the outcome of optimal tax policies that exempt small firms from taxation in order to save on administrative costs. In our model, the missing middle arises not because small firms are tax exempt *de jure*, but because small firms can sustain tax evasion and therefore become tax exempt *de facto*.

Macroeconomic development and optimal government policy

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

Proposition 6. 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 $x(\widehat{N}(A), A) \cdot (1 - \tau(A)) = x(1/\delta, A)$. Firms do not evade taxes. Government size is suboptimal, $\tau(A) < \tau^*$, and $\tau(A)$ is increasing in A.
(3) Late stage: when $A \geq A_H$, the government is no longer constrained by tax enforcement and firms do not evade taxes. The effective tax rate is set at the optimal level $\tau(A) = \tau^*$, and government size is constant in A.

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

$$-\frac{1}{1 - \tau(A)} \frac{\mathrm{d}\tau(A)}{\mathrm{d}A} = \frac{x_A(1/\delta, A)}{x(1/\delta, A)} - \frac{x_A(\widehat{N}(A), A)}{x(\widehat{N}(A), A)}.$$

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

The predictions of Proposition 6 are illustrated in Figure 4. Following an early stage with zero tax revenue and no public goods provision, the government gradually increases the effective 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. It would be straightforward to extend our model to incorporate traditional taxes that do not rely on third-party reporting by assuming that the government can raise a fixed fraction τ_0 of national product through such taxes. In that case, the theoretical path of the tax-to-GDP ratio in Figure 4 would be shifted upwards by τ_0.

We note that the theoretical predictions illustrated in Figure 4 are consistent with the macro Stylized Facts 1 and 2 presented in Section II.

In our theory, in the first two stages of development, the government sets the statutory tax rate at the maximum sustainable level. As a result, there is never any tax evasion and the effective tax rate is always identical to the statutory tax rate. In reality, the effective tax rate and the statutory tax rate differ markedly. Gordon and Li (2009) argue that in contrast to effective tax rates, statutory rates are not systematically different across countries by level of development. Our macro theory cannot capture this phenomenon because there is no heterogeneity across firms at any point in time, which is a limitation of our model. With heterogeneity across firms, large firms would be complying while small firms would be evading. As a result, the statutory and effective tax rates would differ. The effective tax rate would likely still follow an S-shape over the

tax-to-GDP ratio

perfect enforcement

no enforcement
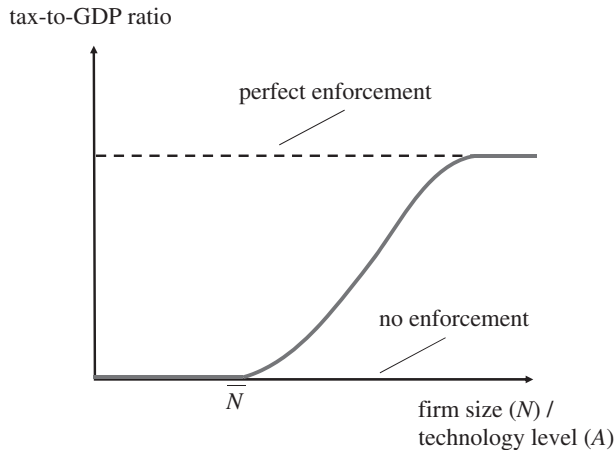
$\bar{N}$

firm size (*N*) /
technology level (*A*)

FIG. 4. Evolution of government size.

process of development (as it starts from zero when all firms are small and ends up at the Samuelson level when all firms are large), but it is not clear what the shape of the statutory tax rate would be. We leave this important extension to future research.

Extensions

Endogenous use of business records Our analysis has assumed the existence of accurate business records ('books') that create potentially detectable evidence of tax evasion. One way for a firm to escape taxation completely is to discard the use of books altogether. As discussed in Section III, being off the books is presumably associated with a productivity loss that is growing in firm size and complexity, and firms choose to be on or off books by trading off this productivity loss against the tax savings as in Gordon and Li (2009). It is conceptually straightforward to set out a macroeconomic model along these lines, which generates results that are fully consistent with those presented above.

We assume that the average product of labour for a firm on the books is equal to $x(N, A)(1-c)$, where c is a fixed administrative cost of maintaining books per unit of output, and x has the same properties as in the earlier model. Average productivity for an off-the-books firm equals $x(N,A)\alpha(N)$, where α reflects the output loss of not having accurate business records. We make the assumptions $\alpha(0) = 1$, $\alpha'(N) < 0$ and $\lim_{N\to\infty} \alpha(N) = 0$.[36] To zoom in on the implications of endogenous books, we drop the agency model of third-party reporting, and assume simply that a firm on the books is perfectly taxable, while a firm off the books cannot be taxed at all. All other components of the model (such as the specification of consumers and government) are exactly as before. In this setup and under some additional regularity assumptions, it is possible to state a result analogous to Proposition 6 and with an evolution of government size as in Figure 4. The mechanism that drives this development is no longer the increased risk of third-party whistleblowing but rather an increased productivity gain of using rigorous business records as firms get larger. The model and results are presented in Subsection B.1 of the online appendix.

Endogenous growth The above analysis of the development of tax enforcement and government size assumes that productivity increases exogenously. This is a reasonable assumption if government activities have only a limited impact on the growth process.

However, some government activities, such as the protection of property rights, law enforcement, and investments in education, health and infrastructure, may be very important for growth. Barro (1990) develops an endogenous growth model where government inputs are complementary to private inputs in production, and derives the optimal effective tax rate and government size along the growth process. It is possible to embed a Barro-type production technology in our theory of tax enforcement, and obtain a model where optimal firm size grows with endogenous technological progress. We present this model in Subsection B.1 of the online appendix. The government applies a time-varying tax rate to maximize the lifetime infinite-horizon utility of a representative household. Under some parameters, an economy might get stuck in a poverty trap, because firms are too small and the government cannot raise income taxes to feed the growth process. When the economy is not stuck in a poverty trap, there will be three stages of development, as in Proposition 6. First, the government cannot raise income taxes and the economy grows too slowly relative to first-best. Second, the government starts raising income taxes, but the effective tax rate is constrained by tax enforcement. The economy grows faster but still slower than first-best. Third, the government is no longer constrained by tax enforcement and can apply the effective tax rate that optimizes the growth rate. Thus this endogenous growth model delivers the same S-shaped time pattern of the tax-to-GDP ratio that fits the empirical evidence. The model also suggests that the inability of some of the poorest countries to start the growth process might be due to insufficient fiscal capacity.[37]

V. Conclusion

We have presented a simple agency model to explain why third-party information 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 enforcement successful is the combination of verifiable book evidence that is common knowledge within the firm, and a large number of employees. As a result, a single employee can denounce collusive tax cheating between employees and the employer by—either accidentally or deliberately—revealing true books to the government.[38] We have embedded this agency model into a macroeconomic growth model where the size and complexity of firms grows with technological progress. Our model is able to capture a set of stylized facts set out at the beginning of the paper, including the S-shaped evolution of the tax-to-GDP ratio driven by the expansion of third-party reported taxes over the course of development. In our model, economic development and the associated change in firm structure relax the tax enforcement constraint and naturally lead to large welfare state governments.

While our theoretical analysis is consistent with the main stylized facts on taxation and development, in future work it would be valuable to directly test the predictions of our model using micro data.

Our theory predicts that third-party enforcement is most successful for large and complex firms. The related theories proposed by Gordon and Li (2009) and Kopczuk and Slemrod (2006) point out that links to the financial sector and the network of arm's-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's-length transactions (as in Kopczuk and Slemrod 2006) produce verifiable information that the government can exploit for tax purposes. Hence it is really the volume of recorded

transactions (both internal and external) that grows with economic development and increases the ability to tax. In principle, an empirical analysis of tax audits of both firms and employees in a developed country could be used to assess which factors—size and complexity, links to the financial sector, network of transactions—explain best the low levels of tax evasion observed in advanced OECD countries.[39]

Our theory also predicts that the inability to collect income taxes from the informal sector is the key reason why developing countries collect little tax revenue.[40] Other theories have been put forward: (1) corruption in the tax administration may make taxes hard to collect in both the formal and informal sectors; (2) demand for government services may be lower in poor countries. We could test our theory by estimating effective tax rates in the formal and informal sectors of developing countries and comparing them with the effective tax rates in OECD countries. Our theory predicts that effective tax rates on the formal sector in developing countries should be high—possibly as high as in OECD countries—while the alternative theories imply that even in the formal sector, effective tax rates should be much lower in developing countries than in OECD countries.

NOTES

1. The withholding system is useful to individuals or companies when there are credit constraints, a point that we will not investigate in this paper, where we focus only on informational aspects.
2. Our model focuses on *internal* information sharing within the firm. However, firms also share information with *external* parties such as other businesses and individual clients, shareholders or debt holders. The number of such external parties also grows with economic development, making tax collusion more difficult, as in our internal information sharing model.
3. 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.
4. As Andreoni *et al.* (1998) conclude at the end of their comprehensive survey (p. 855): 'The most significant discrepancy that has been documented between the standard economic model of compliance and real-world compliance behavior is that the theoretical model greatly over-predicts noncompliance.' Various studies suggest that high compliance rates may be explained by psychological or behavioral aspects such as social norms, tax morale, patriotism, guilt and shame (e.g. Cowell 1990, ch. 6; Andreoni *et al.* 1998, §8). In this paper, we propose instead a theory explaining high compliance based on information.
5. We discuss briefly how the network of firm-to-firm transactions can also help enforcement as firms can also denounce tax cheating of other firms.
6. On the internal side, Crocker and Slemrod (2005) develop a shareholder–manager agency model with tax evasion showing that penalties imposed on managers are more effective in reducing evasion than penalties imposed on shareholders. Chen and Chu (2005) show that the evasion decision of the firm's owner affects the optimal compensation scheme offered to employees and hence creates a distortion in the manager's effort and reduces the efficiency of the contract.
7. See, for example, Alt (1983), Bird (1992), Hinrichs (1966), Kenny and Winer (2006), Webber and Wildavsky (1987). Recently, Sanchez de la Sierra (2015) has proposed an empirical analysis of how local taxation organized by bandits has emerged in the war zones of Eastern Congo in recent years. Such taxation emerges when villages have mineral resources that have good export value and are bulky enough to be fairly easily observable and hence taxable.
8. Note that although value-added taxes do not rely on an explicit system of third-party reporting, they do rely implicitly on third-party information generated by the paper trail between different firms in the value-added chain (see, for example, De Paula and Scheinkman 2010; Pomeranz 2015). Historically, indirect taxes were

applied on goods where the government could monitor transactions, either at the point of production (mining goods such as salt) or at the point of entry (such as tariffs). Enforcement of modern broad-based sales taxes (such as the state-level US sales taxes) relies instead on accounting books of firms and hence are partly modern taxes. We classify them as traditional taxes to be conservative.

9. The case of Sweden is important to show that external shocks and the ensuing ratchet effects are not necessary for the growth of government.

10. Consistent with this finding, Kleven (2014) shows that tax take and the fraction of self-employed workers in the workforce (zero-employee firms) are negatively related across countries.

11. Tax evasion from panel B of Figure 3 is the detected tax evasion, which could be well below the actual tax evasion as tax audits cannot uncover all tax evasion, especially in small informal businesses (see US Internal Revenue Service 2012). Hence it is likely that the actual evasion rates for firms with no employees in Denmark is actually much higher than the 4.7% detected in panel B of Figure 3.

12. Naturally, alternative theories could possibly be also consistent with these three stylized facts. For example, the theory of Dharmapala *et al.* (2011), where the government faces fixed (per-firm) administrative costs of tax collection, predicts Stylized Fact 3 and the first part of Stylized Fact 2, but is silent on Stylized Fact 1 as it does not make the distinction between third-party reported taxes vs. self-reported taxes.

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

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

15. In the final subsection of Section IV, we consider the implications of endogenizing the choice of being on the books as in Gordon and Li (2009).

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

17. To be sure, in practice, profits are different from wages because they are not recorded in the same way. Wages are recorded on payroll lists while profits are typically obtained by substraction as $\Pi = S - P - W$. We further discuss this issue below.

18. For example, in the USA, such reports are made through W2 forms issued by firms and sent to both the government and employees. Employees use this information to file their income tax returns (Logue and Slemrod (2010) discuss this mechanism in detail). Some other OECD countries, such as Denmark, use pre-populated income tax returns whereby the government informs individuals about their earnings using information received from firms.

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

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

21. More precisely, we assume that the outside options for any coalition of individuals are always given by $\overline{y} = (\overline{y}_1, \ldots, \overline{y}_N)$.

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

23. 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.

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

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

26. We discuss in the next subsection whether such a form of whistleblowing rewards can be seen as an optimal mechanism for the government to elicit tax compliance.

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

28. If moral rewards are heterogeneous across individuals and unobservable by the employer, then the model becomes conceptually very close to the random-shock model analysed above.

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

30. As in Gordon and Li (2009), this debt channel is one of the benefits of using accounting books and being formal.

31. No developing country with few large businesses can successfully implement a broad VAT (Ebrill *et al.* 2001). Furthermore, the VAT is not a necessary condition for successful corporate and individual income tax enforcement, as shown by the example of the USA.

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

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

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

35. The empirical finding of a missing middle is debated, however. Hsieh and Olken (2014) find no evidence of a missing middle based on comprehensive manufacturing firm data from India, Indonesia and Mexico.

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

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

38. 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.

39. The recent studies by Pomeranz (2015), Kumler *et al.* (2013) and Carillo *et al.* (2014) provide very compelling empirical evidence in that direction in the context of developing countries.

40. The theory proposed by Gordon and Li (2009) makes the same prediction.

REFERENCES

ACEMOGLU, D. and ROBINSON, J. A. (2000). Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective. *Quarterly Journal of Economics*, **115**(4), 1167–99.

AIDT, T. S. and JENSEN, P. S. (2009). Tax structure, size of government, and the extension of the voting franchise in Western Europe, 1860–1938. *International Tax and Public Finance*, **16**(3), 362–94.

ALLINGHAM, M. G. and SANDMO, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, **1**, 323–38.

ALT, J. (1983). The evolution of tax structures. *Public Choice*, **41**(1), 181–222.

ANDREONI, J., ERARD, B. and FEINSTEIN, J. (1998). Tax compliance. *Journal of Economic Literature*, **36**, 818–60.

AZARIADIS, C. and STACHURSKI, J. (2005). Poverty traps. In P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. **1A**. Amsterdam: North-Holland.

BARRO, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy*, **98**(5), S103–S125.

BAUMOL, W. J. (1967). Macroeconomics of unbalanced growth: the anatomy of urban crisis. *American Economic Review*, **57**(3), 415–26.

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

BECKER, G. S. (1968). Crime and punishment: an economic approach. *Journal of Political Economy*, **76**(2), 169–217.

BESLEY, T. and PERSSON, T. (2009). The origins of state capacity: property rights, taxation and politics. *American Economic Review*, **99**(4), 1218–44.

——— and ——— (2010). State capacity, conflict and development. *Econometrica*, **78**, 1–34.

——— and ——— (2013). Taxation and development. In A. J. Auerbach, R. Chetty, M. S. Feldstein and E. Saez (eds), *Handbook of Public Economics*, Vol. **5**. Amsterdam: North Holland.

——— and ——— (2014). Why do developing countries tax so little? *Journal of Economic Perspectives,* **28**(4), 99–120.

BEST, M., BROCKMEYER, A., KLEVEN, H., SPINNEWIJN, J. and WASEEM, M. (2015). Production vs revenue efficiency with limited tax capacity: theory and evidence from Pakistan. *Journal of Political Economy*, **123**(6), 1311–55.

BIRD, R. M. (1992). *Tax Policy and Economic Development*. Baltimore MD: John Hopkins University Press.

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

CAGÉ, J. and GADENNE, L. (2015). Tax revenues, development, and the fiscal cost of trade liberalization, 1792–2006. Working Paper, Paris School of Economics and University of Warwick.

CARRILLO, P., POMERANZ, D. and SINGHAL, M. (2014). Dodging the taxman: firm misreporting and limits to tax enforcement. NBER Working Paper no. 20624.

CHEN, K.-P. and CYRUS CHU, C. Y. (2005). Internal control versus external manipulation: a model of corporate income tax evasion. *RAND Journal of Economics*, **36**, 151–64.

COWELL, F. (1990). *Cheating the Government*. Cambridge MA: MIT Press.

CROCKER, K. and SLEMROD, J. (2005). Corporate tax evasion with agency costs. *Journal of Public Economics*, **89**, 1593–610.

DE PAULA, A. and SCHEINKMAN, J. A. (2010). Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics*, **2**(4), 195–221.

DHARMAPALA, D., SLEMROD, J. and WILSON, J. D. (2011). Tax policy and the missing middle: optimal tax remittance with firm-level administrative costs. *Journal of Public Economics*, **95**(9–10), 1036–47.

DIXIT, A. K. (2004). *Lawlessness and Economics: Alternative Modes of Governance*. Princeton, NJ: Princeton University Press.

EBRILL, L., KEEN, M., BODIN, J.-P. and SUMMERS, V. (2001). *The Modern VAT*. Washington, DC: International Monetary Fund.

FLORA, P. (1983). *State, Economy, and Society in Western Europe*, 1815–1975, Vol. **1**. London: Macmillan Press.

GADENNE, L. (2014). Nonlinear commodity taxation in developing countries: theory and an application to India. UCL Working Paper.

GORDON, R. and LI, W. (2009). Tax structure in developing countries: many puzzles and a possible explanation. *Journal of Public Economics*, **93**(7–8), 855–66.

HESCH, J. D. (2009). *Reward: Collect Millions for Reporting Tax Evasion. Your Complete Guide to the IRS Whistleblower Reward Program*. Lynchburg, VA: Liberty University Press.

HINRICHS, H. H. (1966). *A General Theory of Tax Structure Change During Economic Development*. Cambridge, MA: Harvard University Press.

HSIEH, C.-T. and OLKEN, B. A. (2014). The Missing 'Missing Middle'. *Journal of Economic Perspectives*, **28**(3), 89–108.

KAU, J. B. and RUBIN, P. H. (1981). The size of government. *Public Choice*, **37**(2), 261–74.

KEEN, M. (2009). VAT, tariffs, and withholding: border taxes and informality in developing countries. *Journal of Public Economics*, **92**(10–11), 1892–906.

KENNY, L. W. and WINER, S. L. (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, H. (2014). How can Scandinavians tax so much? *Journal of Economic Perspectives*, **28**(4), 77–98.

———, KNUDSEN, M., KREINER, C., PEDERSEN, S. and SAEZ, E. (2011). Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark. *Econometrica*, **79**(3), 651–92.

KOPCZUK, W. and SLEMROD, J. (2006). Putting firms into optimal tax theory. *American Economic Review: Papers and Proceedings*, **96**(2), 130–4.

KUMLER, T. J., VERHOOGEN, E. A. and FRIAS, J. (2013). Enlisting workers in monitoring firms: payroll tax compliance in Mexico. NBER Working Paper no. 19385.

LEDERMAN, L. (2010). Reducing information gaps to reduce the tax gap: when is information reporting warranted? *Fordham Law Review,* **78,** 1733–59.

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

LOGUE, K. D. and SLEMROD, J. (2010). Of Coase, Calabresi, and optimal tax policy. *Tax Law Review*, **63**(4), 797–866.

MOORE, J. (1992). Implementation, contracts, and renegotiation in environments with complete information. In J.-J. LAFFONT (ed.), *Advances in Economic Theory: Invited Papers for the Sixth World Congress of the Econometric Society*. Cambridge: Cambridge University Press, pp. 182–281.

MUSGRAVE, R. (1966). *Fiscal Systems*. New Haven CT: Yale University Press.

NARITOMI, J. (2015). Consumers as tax auditors. LSE Working Paper.

OECD (2004). Informal employment and promoting the transition to a salaried economy. In *Employment Outlook*. Paris: OECD, pp. 225–89.

OECD (2006). *Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series*. Paris: OECD.

OECD (2008). *Revenue Statistics: 1965–2007*. Paris: OECD.

PEACOCK, A. and WISEMAN, J. (1961). *The Growth of Public Expenditure in the United Kingdom*. New York: NBER.

PELTZMAN, S. (1980). The growth of government. *Journal of Law and Economics*, **23**(2), 209–87.

POMERANZ, D. (2015). No taxation without information: deterrence and self-enforcement in the value added tax. *American Economic Review*, **105**(8), 2539–69.

Poschke, M. (2014). The firm size distribution across countries and skill-biased change in entrepreneurial technology. IZA Discussion Paper no. 7991.

Riezman, R. and Slemrod, J. (1987). Tariffs and collections costs. *Review of World Economics*, **123**, 545–9.

Sanchez de la Sierra, R. (2015). On the origin of states: stationary bandits and taxation in Eastern Congo. Columbia University Working Paper.

Schoenberg, T. and Voreacos, D. (2012). UBS whistle-blower secures $104 million award from IRS. Bloomberg Business News, 11 September.

SKAT (2009). *Firm Tax Compliance (Virksomhedernes efterlevelse af skattereglerne)*. Copenhagen: Danish Tax Agency.

Slemrod, J. (2004). The economics of corporate tax selfishness. *National Tax Journal*, **57**(4), 877–99.

Surrey, S. S. (1958). Tax administration in underdeveloped countries. *University of Miami Law Review*, **12**, 158–88.

Tybout, J. R. (2000). Manufacturing firms in developing countries: how well do they do, and why? *Journal of Economic Literature*, **38**(1), 11–44.

US Internal Revenue Service (2012). *Tax Gap for Tax Year 2006: Overview*. Washington DC: US Internal Revenue Service.

Webber, C. and Wildavsky, A. B. (1987). *A History of Taxation and Expenditure in the Western World*. New York: Simon and Schuster.

Yaniv, G. (1992). Collaborated employee–employer tax evasion, *Public Finance*, **47**(2), 312–21.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

A Data Description for Figures 1–3, A1
A.1 Cross-sectional data (Figure 1)
A.2 Long time series data in 14 countries (Figures 2 and A1)
A.3 Tax take and tax compliance vs firm size (Figure 3)

B Theoretical Extensions
B.1 Endogenous Use of Business Records
B.2 Endogenous Growth Model

<div align="center">

Online Appendix for

"Why Can Modern Governments Tax So Much?

An Agency Model of Firms as Fiscal Intermediaries"

Henrik Kleven, Claus Kreiner, and Emmanuel Saez

</div>

A Data Description for Figures 1–3, A1

A.1 Cross-sectional data (Figure 1)

The cross country data are gathered in stata format and are available upon request. The list of 72 countries for which we could obtain information include: Albania, Argentina, Armenia, Australia, Austria, Bahrain, Belarus, Belgium, Bhutan, Bolivia, Bosnia and Herzegovina, Bulgaria, Canada, Chile, China, Congo, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Georgia, Germany, Greece, Honduras, Hong Kong, Hungary, Iceland, Iran, Ireland, Israel, Italy, Japan, Kazakhstan, Latvia, Lithuania, Luxembourg, Maldives, Mauritius, Moldova, Mongolia, Morocco, Netherlands, New Zealand, Norway, Paraguay, Peru, Poland, Portugal, Romania, Russia, Seychelles, Singapore, Slovak Republic, South Africa, Spain, St. Kitts and Nevis, St. Vincent and the Grenadines, Swaziland, Sweden, Switzerland, Tajikistan, Thailand, Togo, Ukraine, United Kingdom, United States, Vietnam.

The data are obtained from the OECD (2008) for the 29 OECD countries and the official IMF database for the other 43 countries. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and Value Added Taxes. Traditional taxes include all the other taxes

For non-OECD countries, the IMF lumps together VAT and other taxes on goods (sales and excise taxes) in a single tax on sales category. In that case, we assume that 60% of taxes on sales are from the VAT where 60% is the average VAT to tax on sales ratio among OECD countries (where we have the information).

A.2 Long time series data in 14 countries (Figures 2 and A1)

The time series data are gathered in both excel and stata format with documentation and are available upon request.

Sources: For the period 1965-2008, data comes from the OECD.StatExtracts, Public Sector, Taxation and Market Regulation. Data before 1965 comes from Flora (1983), volume I. In addition, different country specific sources listed below are used to obtain historic values of GDP.

Definition of variables: For the OECD data; we define modern taxes as the sum of taxes from "Income, profits, and capital gains," "Social contributions," "Payroll and workforce" and "VAT". Traditional taxes are then calculated as the residual between total and modern taxes. Taxes include taxes from both central and local governments and hence are a comprehensive measure of the tax burden in each country.

In the data from Flora the decomposing of general taxes is not fully complete and income taxes therefore include the following: (quote from Flora)

"For general government the classification of direct taxes is much simpler. Three categories are distinguished. Income taxes include here, apart from the actual income tax, property, extra-ordinary and corporate taxes."

When defining modern taxes we need to add social contribution to Flora's income taxes to ensure consistency with OECD data. From 1949, Flora presents data for the total expenditure used on social security including social insurance. These expenditures are split up by the financing institution: insured, employers and the public sector. The part paid by the public sector is most likely already accounted for in the calculation of taxes and the part paid by insured persons is presumably voluntary and by definition not a tax. Hence, to avoid including voluntary payments and not to double count, we define social contribution only to be social security paid by employers. We now define modern taxes as the sum of the defined social contributions, income taxes and VAT. Traditional taxes are again simply calculated as the residual.

Interpolations: For all countries there are years with missing data. To deal with this issue; simple linear interpolations have been used between data points. Two special cases are for Social contributions and VAT. Flora presents data for Social contributions from 1949 and onwards. However, these social programs were introduced prior to 1949. For the interpolation, we therefore set the value equal to 0 one year ahead of introducing the program (according to p. 454 in Flora) and interpolate linearly to 1949. For VAT; all countries except France and Finland introduced VAT after 1965 and are thus covered by OECD. For France and Finland we need to add VAT prior to 1965, which is done the same way as with social contributions.

The details regarding sources, data breaks and interpolations are summarized for each country in Appendix table A1. First column shows the country name, the next 4 describe interpolations and the two last summarizes data sources and data breaks. For the interpolations, notation is a bit compact. E.g. for Austria interpolations have been made between 1892-1905, 1905-25 etc. This is then shortened to 1892-1905-25 and so on.

Country specific notes

Austria:
1892: GDP is calculated by interpolating the value from 1913 by assuming same growth as in Germany between 1892 and 1913. Further, the total general tax value from Flora is scaled due to the fact that Flora presents data for the Hapsburg Empire. This is done by using the ratio between the population of Austria in 1890 according to Maddison (1995) and the population of the Hapsburg Empire according to Flora (p. 44).

1905: Same as 1892. Now interpolating the value from 1913 by assuming same growth as in Germany between 1905 and 1913. The taxation value is scaled with the ratio between population of Austria in 1900 according to Maddison (1995) and the population of the Hapsburg Empire according to Flora (p. 44).

Belgium:

1912: In order to calculate total taxation of GDP in 1912; Clark's GDP estimate for 1913 has been combined with Flora's estimate of total taxation in 1912.

1925: In order to calculate total taxation of GDP in 1925; Clark's GDP estimate for 1924 has been combined with Flora's estimate of total taxation in 1925.

Italy:

1949-1965: According to Flora, the total value of social security is calculated in million lire, while total taxation is given in 100,000 lire. However, it seems that the two figures are scaled equally which we assume to be the case.

Switzerland:

1886: In order to calculate total taxation of GDP in 1886; Clark's GDP estimate for 1890 has been combined with Flora's estimate of total taxation in 1886.

1900: In order to calculate total taxation of GDP in 1900; Clark's GDP estimate for 1898 has been combined with Flora's estimate of total taxation in 1900.

A.3   Tax take and tax compliance vs firm size (Figure 3)

Panel A shows the relationship between total the tax to GDP ratio and the share of the workforce working in firms with 10+ employees across 59 countries where information exists. The total tax to GDP ratio is measured in 2012 and is from the Index of Economic Freedom, Heritage Foundation. The share of the workforce working in firms with 10+ employees is computed by combining the Global Entrepreneurship Monitor (GEM) survey data and World Bank data. We first compute for each country the share of employees working in firms with 10+ employees based on the survey question "Not counting the owners, how many people are currently working for this business?" (the variable omnowjob). We compute country-averages based on the survey years 2001-2010. By multiplying the employment share with (1-fraction of workforce self-employed), we obtain the number of employees in 10+ firms in proportion to the workforce. The fraction of self-employed in the workforce is obtained from World Bank data and computed for the latest year possible (more details are in the electronic appendix to Kleven 2014). The GEM survey are conducted by London Business School and Babson College and may be downloaded at http://www.gemconsortium.org.

Panel B shows tax evasion rates among Danish firms by number of employees. Tax evasion is measured in proportion to firm sales revenue. The estimates are based on randomized tax audits of firms implemented in 2007, concerning tax payments of firms in 2006, as part of a large-scale field experiment in Denmark (see Kleven et al. 2011 for an analysis of the individual-level component of this experiment). The data contains audited 1650 firms. More details on the random audits of firms are provided in a report by the Danish Tax Agency (SKAT 2009).

B Theoretical Extensions

B.1 Endogenous Use of Business Records

We assume that all firms have access to the same production technology. For each firm, the average product of labor equals $F(N, A, B)$, where N is the number of employees in the firm, A is a technology parameter that grows exogenously over time, and B is an indicator variable that equals 1 when the firm uses books and equals 0 otherwise. As in Section 5.4.1, we assume

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

where $x(N, A)$ is increasing in A and inversely U-shaped in N (as in Section 5), c is a resource cost in proportion of output of bookkeeping, while $1 - \alpha(N)$ reflects the output loss of not using books. We assume that $\alpha(0) = 1$, $\alpha_N(N) < 0$, and $\lim_{N \to \infty} \alpha(N) = 0$.[41] Let $\hat{N}(A) = \arg\max x(N, A)$. As in Section 5, we assume that technological progress is complementary to labor input, defined as $x_A(N, A) / x(N, A)$ being increasing in N. This implies that $\hat{N}'(A) > 0$. Moreover, we assume that, for all N,

$$\lim_{A \to 0} \hat{N}(A) = 0, \quad \lim_{A \to \infty} \hat{N}(A) = \infty, \quad \lim_{A \to \infty} \frac{x(N, A)}{x(\hat{N}(A), A)} = 0. \tag{7}$$

Under those assumptions, we can prove:

Proposition 7 *There are two cut-off levels $A_L < A_H$ which determine three stages of development:*
(1) Early Stage: when $A \leq A_L$, firms do not use books and the government cannot raise any tax revenue and sets $\tau(A) = 0$.
(2) Intermediate Stage: when $A_L < A < A_H$, the government is constrained by tax enforcement. Firms use books and do not evade taxes. Government size is suboptimal, $\tau(A) < \tau^*$, and $\tau(A)$ is increasing in A.
(3) Late Stage: when $A \geq A_H$, the government is no longer constrained by tax enforcement and firms do not evade taxes. The tax rate is set at the optimal level $\tau(A) = \tau^*$ and government size (relative to total product) is constant in A.

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

$$\pi(N, A, 0) = x(N, A)\alpha(N)N - yN, \tag{8}$$

$$\pi(N, A, 1) = x(N, A)(1 - c)N - \frac{y}{1 - \tau}N, \tag{9}$$

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

---

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

conditions with respect to N equals

$$x_N(N, A) N\alpha(N) + x(N, A)\alpha_N(N) N + x(N, A)\alpha(N) - y = 0 \text{ for } B = 0, \quad (10)$$

$$[x_N(N, A) N + x(N, A)](1 - \tau)(1 - c) - y = 0 \text{ for } B = 1. \quad (11)$$

Let us denote by $N(A, 0)$ and $N(A, 1)$ the optimal choices for N given by (10) and (11). There is free entry/exit of firms and labor is completely mobile across firms. This implies that profits are zero in equilibrium. From the profit expressions (8) and (8), we get

$$y = x(N(A, 0), A) \cdot \alpha(N(A, 0)) \quad \text{for } B = 0, \quad (12)$$

$$y = x(N(A, 1), A)(1 - \tau)(1 - c) \quad \text{for } B = 1. \quad (13)$$

Using these two expressions, the first-order conditions (10) and (11) simplify to

$$x_N(N(A, 0), A)\alpha(N(A, 0)) + x(N(A, 0), A)\alpha_N(N(A, 0)) = 0 \text{ for } B = 0, \quad (14)$$

$$x_N(N(A, 1), A) = 0 \text{ for } B = 1. \quad (15)$$

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

$$N(A, 1) > N(A, 0). \quad (16)$$

Lemma 1 *Our assumption of complementarity implies*

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

Proof of the Lemma: For $B = 1$, we have from (15)

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

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

For $B = 0$, the first-order condition (14) is

$$\Phi(A, N(0)) = x_N(N(0), A)\alpha(N(0)) + x(N(0), A)\alpha_N(N(0)) = 0. \quad (17)$$

At the optimum, we have

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

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

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

From (17), we have

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

where we have used the Envelope Theorem. By inserting (17), we see that

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

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

In equilibrium, firms that offer the highest wages survive. Hence, firms will use books if this implies that they can offer higher wages to the employees. From equations (12) and (13), the condition for using books becomes

$$1 - \tau > \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \cdot \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c}. \tag{18}$$

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

$$1 - \hat{\tau}\left(A\right) = \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \cdot \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c}. \tag{19}$$

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

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

The assumption of complementarity, $x_A\left(N, A\right)/x\left(N, A\right)$ increasing in N, ensures that the constrained tax rate is increasing, $\hat{\tau}'\left(A\right) > 0$. The assumption (7) ensures that

$$\lim_{A \to 0} \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c} = \frac{1}{1 - c} \text{ and } \lim_{A \to \infty} \frac{x\left(N\left(A, 0\right), A\right)}{x\left(N\left(A, 1\right), A\right)} \frac{\alpha\left(N\left(A, 0\right)\right)}{1 - c} = 0, \tag{20}$$

where we have used that $\lim_{A \to 0} \hat{N}\left(A\right) = 0 \Rightarrow \lim_{A \to 0} N\left(A, 1\right) = 0$ implying that $\lim_{A \to 0} N\left(A, 0\right) = 0$ because of (16). In addition, we have used that $\lim_{A \to \infty} \hat{N}\left(A\right) = \infty$ implies $\lim_{A \to \infty} N\left(A, 1\right) = \infty$. Thus, either $\lim_{A \to \infty} N\left(A, 0\right) = \text{constant}$ in which case the assumption $\lim_{A \to \infty} \frac{x(N, A)}{x(\hat{N}(A), A)} = 0$ ensures the last result or $\lim_{A \to \infty} N\left(A, 0\right) = \infty$ in which case $\lim_{N \to \infty} \alpha\left(N\right) = 0$ ensures the last result.

The limits in (20) and the result $\hat{\tau}'\left(A\right) > 0$ imply that there exist A_L and A_H such that the proposition is satisfied. In particular, when $A_L \leq A \leq A_H$, the government sets $\tau(A) = \hat{\tau}(A)$ given by equation (19). QED.

B.2 Endogenous Growth Model

B.2.1 Households

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

$$u = \int_0^\infty \frac{c^{1-\sigma} - 1}{1 - \sigma} e^{-\rho t} dt, \tag{21}$$

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

$$\dot{k} = rk + (1 - t) w - T - c, \tag{22}$$

where k is the capital stock, r is the net-return on savings, w is the pre-tax labor income, t is a tax rate on labor income, while T is a lump sum tax. We assume that the lump sum instrument is restricted $T \leq \beta y$ where β is the maximum fraction of aggregate income y that the government can collect in lump sum taxes. We introduce lump sum taxes so that the government can raise revenue in all stages of economic development as government spending is essential for economic prosperity as we shall see below. Our empirical analysis in Figures 2 and A1 shows indeed that governments were able to raise a modest fraction of GDP in taxes before modern income and value added taxes became enforceable. Maximization of (21) subject to (22) and a no-Ponzi game condition gives the standard Keynes-Ramsey rule

$$\gamma_c \equiv \frac{\dot{c}}{c} = \frac{r - \rho}{\sigma}. \tag{23}$$

B.2.2 Firms and Productivity

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

$$y_i = f(n_i, k_i, g, k) = x\left(\frac{n_i}{\hat{n}(k)}\right) g^\alpha k_i^{1-\alpha} n_i^\alpha, \tag{24}$$

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

Moreover, we assume that capital and firm size are complementary, $\hat{n}'(k) > 0$, reflecting that the workforce needs to organize in larger firms in order to reap the full return of a larger capital stock. Importantly, we assume for simplicity of exposition that the optimal firm's size depends on the aggregate capital stock k and not on the firm's specific capital stock k_i. Finally, note that the capital stock of each firm is negligible compared to the aggregate capital stock as there is a continuum of firms. In a symmetric equilibrium, each firm employs n workers. There is therefore a continuum of firms of measure $1/n$ (as there is a continuum of workers of measure one). Each firm employs $k_i = n \cdot k$ units of capital where k is the aggregate capital stock. Hence, summing (24) across all identical $1/n$ firms, aggregate production is

$$y \equiv x\left(\frac{n}{\hat{n}(k)}\right) g^\alpha k^{1-\alpha} \leq g^\alpha k^{1-\alpha}, \tag{25}$$

7

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

Profits of firm i equal

$$\pi_i = x\left(\frac{n_i}{\hat{n}(k)}\right) g^\alpha k_i^{1-\alpha} n_i^\alpha - r k_i - w n_i. \tag{26}$$

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

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

which gives

$$
\begin{aligned}
w &= \left[\frac{x'(\cdot)}{x(\cdot)} \frac{n_i}{\hat{n}} + \alpha\right] x(\cdot) g^\alpha k_i^{1-\alpha} n_i^{\alpha-1}, \tag{27} \\
r &= (1-\alpha) x(\cdot) g^\alpha k_i^{-\alpha} n_i^\alpha. \tag{28}
\end{aligned}
$$

From equations (26)–(28), we obtain

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

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

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

$$
\begin{aligned}
y &= g^\alpha k^{1-\alpha}, \tag{29} \\
w &= \alpha g^\alpha k^{1-\alpha}, \tag{30} \\
r &= (1-\alpha) g^\alpha k^{-\alpha} \tag{31}
\end{aligned}
$$

where we have used equations (25), (27), (28), and $n_i = \hat{n}(k)$. Note that the standard macroeconomic equation $y = w + rk$ holds.

B.2.3 Optimal Government Policy

Case with No Tax Evasion

We consider a benevolent government that chooses (g, T, t) in order to maximize lifetime utility (21). The government policy has to satisfy the government budget constraint

$$g \le T + tw. \tag{32}$$

Let $\tau \equiv g/y$ denote the government to GDP ratio. From equation (29), we have

$$g/k = \tau^{\frac{1}{1-\alpha}} \tag{33}$$

From equations (22), and (32) and (33), we obtain the resource constraint

$$\dot{k} = g^\alpha k^{1-\alpha} - g - c = \tau^{\frac{\alpha}{1-\alpha}} k - g - c. \tag{34}$$

8

From equations (23) and (26), we obtain the steady state growth rate of consumption for a given government to GDP ratio τ

$$\gamma_c = \frac{(1-\alpha)(g/k)^\alpha - \rho}{\sigma} = \frac{(1-\alpha)\tau^{\frac{\alpha}{1-\alpha}} - \rho}{\sigma}, \tag{35}$$

which also becomes the steady state growth rate of k and y. The marginal benefit of raising public spending is $\alpha g^{\alpha-1} k^{1-\alpha} = \alpha (g/k)^{\alpha-1}$ while the marginal cost is 1. This, together with equation (33), implies that the optimal policy solution that decentralizes the first best allocation is $\tau^* = \alpha$ in which case, the growth rate of the economy becomes

$$\gamma_c = \frac{(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} - \rho}{\sigma}, \tag{36}$$

which is constant over time.[42]

Case with Full Tax Evasion

With full tax evasion, it is impossible to tax income, $t = 0$. We now have $g = T \leq \beta y$. We assume $\beta < \alpha$ implying that it is impossible to attain the optimal government-to-GDP ratio with lump sum taxation alone. From (35), we obtain the growth rate

$$\gamma_c = \frac{(1-\alpha)\beta^{\frac{\alpha}{1-\alpha}} - \rho}{\sigma}. \tag{37}$$

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

Case with Tax Enforcement

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

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

Proposition 8 *Let $\bar{n} \equiv 1/\delta$. We obtain the following cases:*

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

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

---

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

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

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

Proof: Without tax evasion, $n_i = \hat{n}(k)$ is optimal and the pre-tax wage rate as a function of the capital stock is given by equation (30) such that $w = \alpha \cdot g \cdot (k/g)^{1-\alpha}$. The capital stock as a function of the real interest rate from equation (31) is such that $k = \left(\frac{1-\alpha}{r}\right)^{\frac{1}{\alpha}} g$. By inserting this expression, in equation (30), we obtain

$$\bar{y} = (1-t)\, w = (1-t)\,\alpha g \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}}. \tag{38}$$

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

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

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

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

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

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

The company has to offer each worker at least $\bar{y}$ in (38). This implies

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

which is fulfilled if

$$t \leq 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}. \tag{39}$$

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

The proof of (1) follows from the fact that profits are always greater under evasion when this can be sustained at the optimal firm size $\hat{n}(k)$. The proof of (2a) and (2b) follows directly from the above argument and the condition (39). QED.

Macroeconomic Development

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

10

Proposition 9 *Optimal government policy and possible stages of economic development*

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

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

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

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

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

Proof: In all cases, the economy starts with $k < \underline{k}$ so that firms are untaxable initially. Suppose that $\beta \leq \left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$, then equation (37) implies that the growth rate is negative. In that case, the economy is stuck in a poverty trap which proves (1).

Suppose instead that $\beta > \left(\frac{\rho}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}}$, then equation (37) implies that the growth rate is positive. As $\beta < \alpha$, the growth rate is lower than the first best growth rate given by (36) which proves (2a).

As the economy has a positive growth rate, k will eventually reach $\underline{k}$ and Proposition 8, (2b) implies that a maximum tax at rate $t = 1 - x\left(\frac{\bar{n}}{\hat{n}(k)}\right)^{\frac{1}{\alpha}}$ can be enforced, which proves (2b).

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

Additional References

Clark, Colin. (1957), *The Conditions of Economic Progress,* Macmillan & Co LTD: London.

Flora, Peter. (1983), *State, Economy and Society in Western Europe 1815-1975,* Campus Verlag: Frankfurt.

Hjerppe, Riitta. (1989), *The Finnish Economy,* Bank of Finland / Government Printing Centre: Helsinki.

Maddison, Angus. (1995). *Monitoring the World Economy, 1820-1992*, OECD Publishing: Paris.

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

Osterreichs Volkesinkommen 1913 bis 1963, (1965), Monatsberichte des Osterreichischen Institutes fur Wirtschaftsforschung.

Toutain, Jean-Claude. (1987), "Le Produit Intérieur Brut de la France de 1789 à 1982," Cahiers de l'I.S.M.E.A. Série Histoire Quantitive de l'Économie Française, no. 15.

Figure A1: Tax Take and Tax Structure Over Time
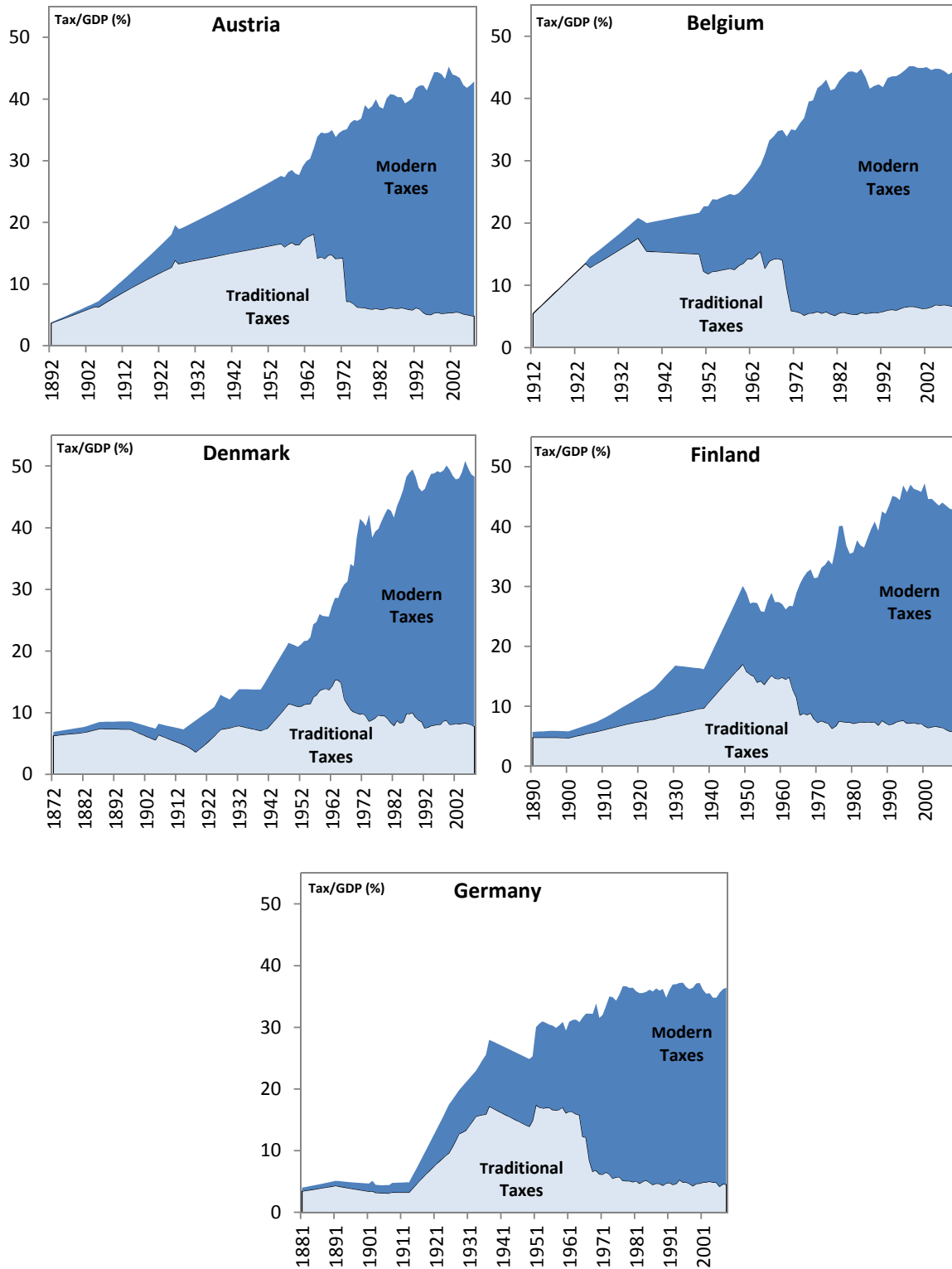
Notes and sources: Data from Flora (1983) before 1965 and from OECD after 1965. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and value-added taxes. Traditional taxes include all the other taxes. See appendix for complete details.

Figure A1: Tax Take and Tax Structure Over Time (cont.)
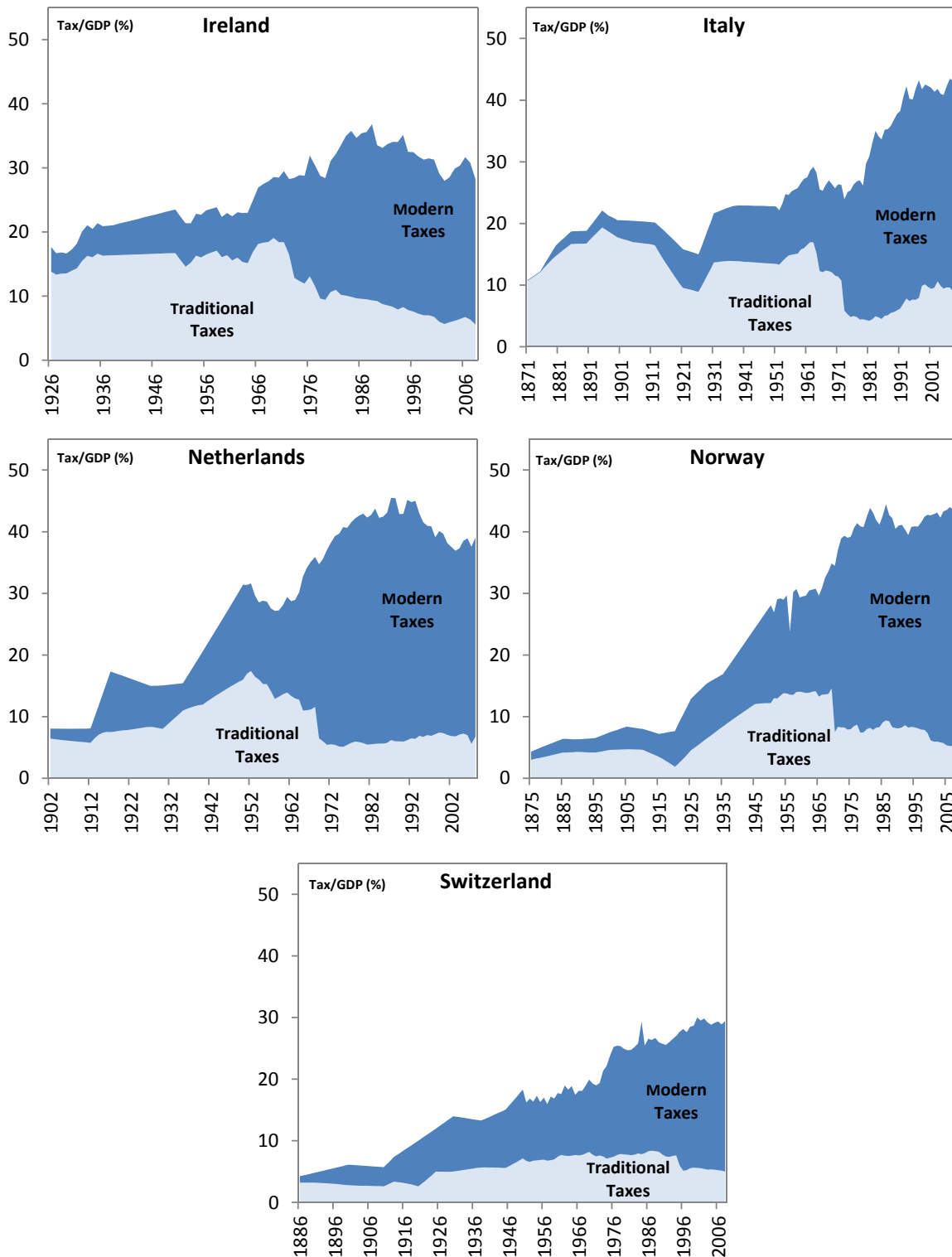
Notes and sources: Data from Flora (1983) before 1965 and from OECD after 1965. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and value-added taxes. Traditional taxes include all the other taxes. See appendix for complete details.

Appendix Table A1: Interpolations in Creating the Time Series Data

	Interpolations				Sources and time span	Data breaks
	Total taxes of GDP	**Income taxes of total taxes**	**Social Cont. of total taxes**	**VAT of total taxes**		
AU	*1892-1905-25-28-55*	**1905**-25-28-55	**1886**-1955	-	1892-1925: Flora and Öst. Vol. 1892-1965: Flora 1965-2008: OECD	1924-25 1964-65
BE	*1912-25-36-38-50* *1954-57* *1962-65*	**1925**-36-38-50 **1954**-57 **1962**-65	**1923**-51 **1954**-57 **1962**-65	-	1912-36: Flora and Clark 1910-65: Flora 1965-2008: OECD	1935-26 1964-65
DE	*1872-82-87-97* *1906-14-24-26-29-32-39-48-51*	**1872**-82-87-97 **1906**-14-18-24-26-29-32-39-41-48-51	**1915**-51	-	1872-1965: Flora 1965-2008: OECD	1964-65
FI	*1890-96-1900-08-10-24-30-38-49*	**1890**-96-**1900**-08-10-24-30-38-49	**1894**-1949	**1963**-65	1890-1930: Flora and Hjerppe 1890-1965: Flora 1965-2008: OECD	1929-30 1964-65
FR	*1850-55-60-65-70-75-80-85-90-95-1900-05-10-20-25-30-35-52* *1963-65*	**1916**-20-25-30-35-40-45-50 **1963**-65	**1909**-50 **1963**-65 **1966**-68	**1953**-65	1850-1920: Flora and Toutain 1850-1965: Flora 1965-2008: OECD	1919-20 1964-65
GE	*1881-91-1901* *1903-05-07* *1908-13-25-28-30-33-36* *1937-50*	**1881**-91-1901 **1903**-05-07 **1908**-13-25-28-30-33-36 **1937**-50	**1882**-1949	-	1881-1965: Flora 1965-2008: OECD	1964-65
IR	*1936-38-50-52* *1956-58*	*1936-38-50-52* *1956-58*	**1910**-50-52 **1956**-58 **1967**-71	-	1926-65: Flora 1965-2008: OECD	1964-65
IT	*1871-74-75-80-85-90-95-1900-12-21-26-31-36-38* *1939-51*	*1871-74-75-80-85-90-95-**1900**-11* *1912-21-26-31-36-38* *1939-51*	**1897**-1951	-	1871-1965: Flora 1965-2008: OECD	1964-65
NE	*1902-07-12-17-22-27-30-35-51* *1954-56* *1959-61*	**1893**-97-1902-07-12-17-22-27-30-35-40-50	**1900**-50	-	1893-1965: Flora 1965-2008: OECD	1964-65
NO	*1875-80-85-90-95-1900-05-10-15-20-25-30-35-50* *1959-61*	*1875-80-85-90-95-1900-05-10-15-20-25-30-35-40-45-50* *1959-61*	**1893**-1950 **1959**-61	-	1875-1965: Flora 1965-2008: OECD	1964-65
SW	*1880-90-1900-10-13-18-23-28-33-38-43-49*	*1910-13-18-23-28-33-38-43-49*	**1912**-49	-	1880-1965: Flora 1965-2008: OECD	1964-65
SZ	*1886-1900-10-13-30-38-45-50*	*1886-1900-10-13-20-25-30-38-45-50*	**1910**-50	-	1886-1910: Flora and Clark 1886-1965: Flora 1965-2008: OECD	1909-10 1964-65
UK	*1868-72-75-80-85-90-95-1900-05-10-15-20-25-30-35-40-45-50*	*1868-72-75-80-85-90-95-1900-05-10-15-20-25-30-35-40-45-50*	**1910**-50	-	1868-1965: Flora 1965-2008: OECD	1964-65