CENTRE DE RECHERCHES EN ÉCONOMIE ET DROIT

Paris Center for Law and Economics

CRED WORKING PAPER N^o 2019-13

The Transition from Tax farming to Public Tax Collection: from Qaids and Pashas to Civil Servants

Touria Jaaidane*

*LEM-CNRS (UMR 9221), Université de Lille, Faculté des Sciences Economiques et Sociales. Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France, and CRED - Université Paris II. Email : touria.jaaidane@univ-lille.fr.

Abstract

We propose a model to explain the transition from privately to publicly organized tax collection. We consider a relationship between a central government and a local agent in charge of collecting taxes and enforcing the law. We compare delegation to a private collector, *i.e.* a tax-farmer, by means of a *standard debt contract* à la Townsend (1979), and to a public agent, thanks to an incentive contract. Both arrangements leave the agent with a rent. The transition to public collection is driven by state verification costs, the public agent's monitoring costs and limited liability constraint. To assess the relevance of our analysis for the interpretation of the History of Institutions, we consider the move of Great-Britain's and France's from the General Farm to a centralized public tax collection system and the pre-protectorate Morocco move from reliance on Qaids, as private tax collectors, to civil servants, when the country became independent in 1956.

Keywords: tax collection; tax-farming; law enforcement; Principal-Agent contract; incentives; costly state verification model; standard debt contract.

JEL Classification: H11, K42, N40.

1 Introduction

Public finance institutions, and, more specifically, tax collection systems are essential to the building of states. The government's ability to borrow, spend and grow, as well as the political stability of nations, crucially depend on the organization of tax collection. The analysis of the causes and consequences of taxation is a common research concern to economists, historians, political scientists and sociologists (see Matthews (1958), Tilly (1975), Mathias and O'Brien (1976) and O'Brien (1988)). The study of pre-modern (from 1500 to 1800) European fiscal regimes has led to an extensive literature (see Bonney (1995)) trying to understand how rulers were able to tax their subjects. The contributions of Coşgel (2005) offer insightful explanations on the building of the fiscal regime of the Ottoman Empire.

As documented by historians, various forms of delegated tax collection did exist in the past, relying on rent contracts, revenue-sharing contracts or plain wage contracts. Under the rent contract, or tax farming contract, the agent pays a fixed rent to the King, and keeps the tax revenues, net of the rent, as an income. With a sharing contract, the agent is leased the right to collect taxes and receives a share of the revenue. With a plain wage contract, the agent receives a fixed compensation for his work, and returns all the tax revenues to the government.

The present work is about the transition from private tax collection to publicly organized fiscal administration. We analyze a simple model where a government, hereafter called the King, delegates both the tasks of tax collection and law enforcement to an agent. This agent collects taxes and exerts a costly effort to detect fraud. The government can contract with either a private or a public agent, *i.e.*, a tax farmer or a civil servant. In both cases, the agent has private information about a random shock affecting the local economic conditions.

We make different assumptions on how well the central government is informed about the collected tax revenues. We first assume that the arrangement with the private agent is based on a *standard debt contract*, derived within the Costly State Verification Model (hereafter, CSV), pioneered by Townsend (1979) and Gale and Hellwig (1985). The basic feature of a standard debt contract (hereafter, SDC) is the borrower's commitment to a constant repayment across states of nature, and the lender's right to seize the entire cash flow if this repayment requirement cannot be met. We view the CSV (and the SDC) as the appropriate model for tax farming arrangements in which a private agent pays a lease price in exchange for the right to collect taxes and to enforce the tax law. In the CSV setting, the King (*i.e.*, the Principal) does not observe the amount of tax revenues, but can learn this amount by means of a costly audit. Yet, the King never observes the level of effort exerted by the agent. The verification costs are incurred only in the event of the Agent being unable to pay the lease, in which case the Principal confiscates the entire revenues. The observation costs borne by the King may be interpreted as expenses, made to wage military or police expeditions, aimed at revealing the amount of collected taxes.

Under the alternative contract, the principal offers an employment contract to a public agent, who we call a civil servant. The contract specifies a compensation consisting of a fixed part and a variable part, based on a performance measure. In this case, the principal observes the revenues by means of some accounting system, but observes neither the realization of the random local shock, nor the agent's effort. The resources devoted to building an accounting and reporting system are considered as a fixed cost of this contract with a public agent. This public organization regime can be modeled by means of Principal-Agent theory. The contract will be called hereafter the PAC. We assume that the civil servant is risk-neutral but has a limited liability. In other words, there is a limit to the magnitude of the penalties that can be inflicted in the event of a fiscal shortfall. However, our setting differs from the classical Principal-Agent model à la Grossman and Hart (1983), in which the Agent chooses the effort variable without knowing the realization of a random shock, which is beyond his control. We adopt a setting reminiscent of the approach of Baker (1992) and Baker, Gibbons and Murphy (1994) where the Agent knows when it is worthwhile to work hard: more precisely, in these models, the Agent chooses the effort level after observing a private signal. If the performance measure of the Principal is sufficiently precise and if the parties to the contract have sufficiently congruent objectives (in our context, they are both better off with large tax revenues), then the Principal should propose an incentive contract granting a percentage of the income to the agent.

Tax farming seems to be the efficient mode of organization when "state observation", *i.e.*, the observation of tax revenues, is very difficult, and can clearly be modeled by means of an SDC. In contrast, the delegation of tax collection to a public agent thanks to a contract specifying reward as a percentage of the collected taxes seems relevant if the information and accounting systems are sufficiently developed. Agent compensation can then be based on a reasonably reliable performance measure. Principal-Agent theory then predicts that the choice between the different contracts depends on parameters such as the probability distribution of tax revenues, the risk aversion of both parties, the ability of the State to borrow, the tax collection technology, and the Agent monitoring technology.

To sum up, when the monitoring costs are prohibitive, an SDC is more appropriate; when both are risk-averse, and a performance measure is available, a sharing contract should be chosen; finally, when a risk-neutral principal faces an infinitely risk-averse agent, and the monitoring costs are small, a plain wage contract should be preferred.

We first consider the ideal case of no delegation and study the policies that would be chosen respectively by a benevolent planner and by a profit-maximizing King. The benevolent planner chooses a fraud detection effort allowing for the internalization of the impact on the tax base (*i.e.*, such that the increase in the enforcement costs is just compensated by the change in the collected taxes). We show that the King exerts a larger fraud detection effort and sets a heavier tax rate.

In the study of the two delegation arrangements, we first show that the private tax collector chooses the same policy as the profit-maximizing King. Under an SDC, the risk-neutral private agent becomes the residual claimant for the total fiscal revenues. It follows that the Agent has the proper incentives and replicates the King's centralized policy. Depending on the value taken by the verification costs and the Agent's reservation profit, the solution can be described by two regimes: in the first, the Agent's participation constraint is binding, while in the second, the Agent is left with a rent.

We show that the second-best effort level chosen by the civil servant is smaller than the first-best level that would result if the King observed the effort perfectly. The civil servant also enjoys a rent because his limited liability constraint prevents the King from inflicting high penalties. The second-best tax level is smaller than its first-best level. The distortions of the effort and tax rate meet the objective of reducing the civil servant's information rent. Finally, we show that there are regions in parameter space where the private agent with an SDC is more profitable than the public agent with a wage contract, depending on the values of the respective monitoring costs and the stringency of the limited liability constraint.

Our work is related to the literature on optimal law enforcement (Garoupa (1997)) following the approach that Becker (1968) has applied to criminal law. However, little use is made of Principal-Agent modeling in the economic literature on law enforcement, with the exception of Becker and Stigler (1974), who studied the prevention of corruption; and, more recently, Dharmapala *et al.* (2015) who model an agency relationship endowing the law enforcers with specific preferences (punitive bias). We do not assume that the Agent has such intrinsic motivations; our Agent is simply assumed to be motivated by profit. The comparison between the benevolent planner's and the King's profit-maximizing policies reveals a result in line with the literature on private law enforcement[1]. Over-deterrence is possible here because the private tax collector is the sole law enforcer. In contrast, in Landes and Posner (1975), over-deterrence is the result of competition in the market for law enforcement. The choice between delegation to a private or a public agent has also been studied in the incomplete contracts literature; see *e.g.*, among other contributions, Hart *et al.* (1997)[2]. As we will see, avoiding over-deterrence and over-taxation gives further reasons for public provision of the two tasks[3]. Finally, there is another strand of literature on tax collection (see Coşgel *et al.* (2011)), to which we refer. Notably, Toma and Toma (1992), Kiser (1994) and White (2004) insist on the importance of monitoring costs in the choice between delegating and centralizing tax collection. Priks (2005) shows that this choice depends on the trade-off between ex-ante inefficiency due to the auction mechanism behind the tax farming, and the monitoring cost of the agents under the direct tax collection system. He shows that, when monitoring costs are large, tax farming is preferred, but this leaves a potentially large rent to the tax collector. More recently, Johnson and Koyama (2014) proposed an analysis of the shift from the competitive bidding regime used both in England and France, to a regime they call cabal tax farming, a consolidation of tax farmers into a single entity. This provided public authorities with increased access to short run lending and provided them with the incentives to invest in fiscal capacity (for instance, more standardization of weights and measures and harmonization of the legal environment).

Our setting differs from the above contributions in that we do not model the auction mechanism, but rather focus on the SDC arrangement and the Principal-Agent contract to explore their properties. Moreover, we explicitly model the private tax collector as also being a law enforcer. To sum up, the contribution of the present paper is to model the shift from tax farming

[1]See Friedman (1984), Friedman (1999), Garoupa and Klerman (2002).

[2]See also Donahue (1989).

[3]For a complete discussion of the case for public law enforcement, see Polinsky and Shavell (2007).

to public tax collection as a choice between an SDC à la Townsend and a standard Principal-Agent contract,that we shall call PAC, driven by an improvement in the informational conditions.

A discussion is then developed to confront our model with historical facts. Our model does not reproduce all the varieties of observed fiscal regimes across time and space, yet it helps understand why some states chose a privatized tax collection by entrepreneurs by means of SDCs while others opted for bureaucracies of officials hired under standard PAC. Our analysis is rich enough to encompass several factors explaining the choice between the two regimes, illustrated by historical facts.

The driving forces that are identified are the incentive properties of SDCs, leading to efficient tax-collection technology; the costs K borne by the ruler to verify the tax revenues; the effort-monitoring costs of tax collectors C; the amount R necessary to protect collectors against variability in tax revenues; and, finally, the harmful distortions of SDCs. We will examine the case of Great-Britain's and France's shift from General Farm tax collection to a centralized public system and the pre-protectorate Morocco transition from the *qaids* system of private tax collection to civil servants and modern fiscal administration when the state recovered its independence in 1956.

The choice of England and France, is mainly due to the prominent position of these two states which have built empires; a secondary reason is the existence of a large literature and records on this period. They share a common path in terms of development of fiscal regimes but not at the same period of history. The choice of the Moroccan experience illustrates the evolution in a very different category of country and the influence of colonial rule.

Section 2 presents the model underlying our analysis. In Section 3, we study the case of centralized tax collection, and compare the features of the optimal policies chosen by a benevolent planner and the King. Section 4 derives the optimal policy under the two alternative arrangements described above and studies the theoretical determinants of a transition from tax farming to public tax collection. In section 5 we assess the relevance of our analysis to support the historical evidence.

2 The model

Citizens have a constant productivity θ. The King provides a public good: the protection of property rights, financed out of taxes denoted t. An individual who pays the tax enjoys the net surplus $u_c = \theta - t$, and otherwise, bears the risk of being robbed and of suffering an expected loss x. The expected loss x is a random variable and its distribution depends itself on a shock ϵ, describing the more or less secure local environment. Functions $g(x|\epsilon)$ and $G(x|\epsilon)$ respectively denote the probability density and cumulative distribution functions of x conditional on ϵ. In fact, variable x can be interpreted as an individual's *type* in state ϵ, and the distribution of types may vary with ϵ. In addition, ϵ is randomly distributed on the interval $[\epsilon_0, \epsilon_1]$, with $\epsilon_0 > 0$.

Let $f(\epsilon)$ and $F(\epsilon)$ respectively denote the pdf and cdf of ϵ. Tax fraud is detected with a probability p and the citizen is charged a fine s if caught, so that, for an individual characterized by x, the corresponding net expected surplus of fraud is $u_n = \theta - x - ps$. We assume that $s < \theta$. We assume $x \geq 0$, in other words, an individual is better protected when he(she) pays the tax. A subject pays the tax if $u_c \geq u_n$, that is, equivalently, if $x \geq \tilde{x}$, where by definition,

$$\tilde{x} = t - ps. \tag{1}$$

The tax base is simply the number of compliers, $1 - G(\tilde{x}|\epsilon)$. It is increasing with the expected fine ps and decreasing with the tax t. The King's total revenue from taxes and fines is given by,

$$T(\epsilon) = psG(\tilde{x}|\epsilon) + t(1 - G(\tilde{x}|\epsilon)). \tag{2}$$

To lighten notation, we use $T(\epsilon)$ instead of $T(p, t, \epsilon)$, but T is clearly a function of policy variables (p, t).

By definition, the detection effort p belongs to $[0, 1]$ and it is natural to assume that t belongs to $[0, \theta]$.

To simplify notation again, when necessary, the partial derivative of any function h with respect to any variable z will be denoted,

$$h_z = \frac{\partial h}{\partial z}, \quad \text{and} \quad h_{zz} = \frac{\partial^2 h}{\partial z^2}$$

denotes the second-order partial derivative.

Assumption 1. We assume that x is distributed over the interval $[0, b(\epsilon)]$ with a positive density $g(x|\epsilon)$. The cdf G is twice continuously differentiable with respect to x and ϵ, and $G_\epsilon(\tilde{x}|\epsilon) > 0$ for all x.

When ϵ varies, the whole distribution of losses is potentially modified. This is as if the the distribution of taxpayers' types varied with ϵ. The larger the value of ϵ, the stronger are the incentives for tax evasion.

Assumption 2. The density f is positive on $[\epsilon_0, \epsilon_1]$ and $F(\epsilon)/f(\epsilon)$ is an increasing function[4] of ϵ.

Assumption 3. The conditional density

$$h(x|\epsilon) = \frac{g(x|\epsilon)}{1 - G(x|\epsilon)}$$

is non-decreasing[5] with respect to x.

Assumption 4. We have $2g(x|\epsilon) + xg_x(x|\epsilon) > 0$ at any point x such that $xh(x|\epsilon) = 1$.

Assumption 4 is satisfied, for instance, when g is exponential, or uniform. Finally, we assume that the enforcement costs $\psi(p)$ are an increasing and convex function of p. It follows that p can be interpreted as an effort variable. For simplicity, we will take

$$\psi(p) = \frac{\delta}{2}p^2.$$

3 Centralized tax collection

3.1 The benevolent planner's policy

The benevolent utilitarian planner provides people with protection, collects the tax and enforces the law. The monetary fine s is exogenous, and therefore, for simplicity, fixed. We assume that the planner observes the shock ϵ. It follows that the decisions, denoted $p(\epsilon)$ and $t(\epsilon)$, are functions of ϵ and maximize the following welfare function for all ϵ.

$$W(\epsilon) = \int_0^{t-ps} (\theta - x - ps)g(x|\epsilon)dx + \int_{t-ps}^{b(\epsilon)} (\theta - t)g(x|\epsilon)dx, \qquad (3)$$

[4] The monotone hazard condition is satisfied by most usual distributions: uniform, exponential, normal, logistic, chi-squared, and Laplace.

[5] $h(\hat{x}|\epsilon)$ is the probability that the expected loss x belongs to $[\hat{x}, \hat{x} + dx]$ knowing that $x \geq \hat{x}$.

under the following budget constraint, that should be satisfied for all ϵ. Denoting again $\tilde{x} = t - ps$, we must have,

$$psG(\tilde{x}|\epsilon) + t(1 - G(\tilde{x}|\epsilon)) = \psi(p) \tag{4}$$

We can state the following proposition,

Proposition 1. *If the benevolent planner observes ϵ, the optimal policy $(p^o(\epsilon), t^o(\epsilon))$ satisfies the budget constraint (4) and is characterized by the following properties,*

$$s\eta(\tilde{x}^o(\epsilon) \mid \epsilon) = \psi'^o(\epsilon)), \tag{5}$$

$$0 < \eta(\tilde{x}^o(\epsilon) \mid \epsilon) < 1, \tag{6}$$

where

$$\tilde{x}^o(\epsilon) = t^o(\epsilon) - p^o(\epsilon)s$$

and

$$\eta(x|\epsilon) = \frac{xg\,(x \mid \epsilon)}{1 - G\,(x \mid \epsilon)} \tag{7}$$

is the elasticity[6] of the tax base with respect to x.

For a proof, see the appendix.

Condition (5) determines the optimal fraud detection probability. This condition shows that changing p affects the threshold $\tilde{x}$, and hence the tax base, the fine base and the enforcement costs. When p increases such that the threshold $\tilde{x}$ decreases by 1%, the mass of compliers increases by $\eta(\tilde{x}^o|\epsilon)$%. The marginal cost of enforcement is $\psi'(p)$ and the marginal gain of effort is $\eta(\tilde{x}^o|\epsilon)s$. The optimal fraud detection probability equates the marginal gain of an increased detection probability and the marginal enforcement cost. Condition (6) is a consequence of optimality subject to the budget constraint.

3.2 The King's profit-maximizing policy

Assume for a moment that the King can observe ϵ. For each ϵ, the King would then choose $p(\epsilon)$ and $t(\epsilon)$ to maximize the profit objective,

$$V(\epsilon) = T(\epsilon) - \psi(p). \tag{8}$$

[6]By Assumption 3, it follows that $\eta(x|\epsilon)$ is increasing with respect to x.

We can state the following result.

Proposition 2. *When ϵ is observed, the optimal policy, $(p^*(\epsilon), t^*(\epsilon))$ is characterized by the following properties. The optimal detection probability $p^*(\epsilon)$ is a constant, denoted p^* and for all ϵ, we have,*

$$s = \psi'(p^*); \tag{9}$$

$$\eta(\tilde{x}^*|\epsilon) = 1; \tag{10}$$

$$p^* > p^o \qquad \text{and} \qquad t^* > t^o, \tag{11}$$

where $\tilde{x}^*(\epsilon) = t^*(\epsilon) - p^* s$.

For a proof, see the appendix.

Condition (9) equates the marginal gain and the marginal cost of fraud detection. This determines a constant p^*. In contrast, condition (10) implies that $t^*(\epsilon)$ depends on ϵ. The benevolent planner optimally sets the probability of fraud detection, adjusting the tax rate to internalize the impact on the budget constraint. In contrast, the King simply maximizes profit: his policy entails over-deterrence and over-taxation, as compared to the benevolent planner's policy, as shown by (11).

4 Delegated tax collection

Tax collection and law enforcement are often delegated. The Agent is endowed with a taxation technology and observes the relevant information ϵ. In our model, delegation to an agent is essentially due to asymmetric information, but may also be due to a greater efficiency of the agent in the taxation and enforcement activities. We now turn to the analysis of the two alternative contracts, under different observability assumptions.

4.1 Delegating to a private agent under an SDC

We adapt the costly state verification framework of Townsend (1979) to our analysis. On this approach, see also Gale and Hellwig (1985) and Tirole (2006, Chap. 3, Section 7). Townsend considers a model with a continuum of possible values for the return on the borrower's investment project and the audit policy is deterministic (*i.e.*, the probability of an audit is constrained to take values in the set $\{0, 1\}$). It can be proved that a standard debt contract (SDC) can do as well as any optimal contract. An SDC has the following structure: the borrower must repay a fixed sum D; if the borrower's income is smaller than D, then the lender observes the state of nature: an audit is triggered, the audit cost is paid by the lender, and the lender seizes the entire income. The event in which an audit is triggered can be interpreted as bankruptcy[7]. We assume that the Agent's effort p is never observed by the King. We also assume that the tax revenue T of the Agent cannot be observed by the King, unless he decides to observe the state. The agent pays a lease D in exchange for the right to collect taxes and to enforce the law in his local territory, and earns the income $T - D - \psi(p)$, if $T - D - \psi(p) \geq 0$. The King can learn the net revenue $T - \psi(p)$, provided that he pays the state observation cost, denoted K. This cost is incurred only if there is an audit. Under an SDC, an audit takes place when $T - \psi(p) < D$, in which case the King earns $T - \psi(p) - K$.

In the following, for any function $\phi(\epsilon)$, the expectation of ϕ is denoted

$$\mathbb{E}(\phi) = \int_{\epsilon_0}^{\epsilon_1} \phi(\epsilon) f(\epsilon) d\epsilon.$$

4.1.1 The private agent's optimal policy under an SDC

The Agent privately observes ϵ and chooses $t(\epsilon)$ and $p(\epsilon)$ in order to maximize the fiscal revenues, net of the lease and enforcement costs, that is, to maximize $T(\epsilon) - D - \psi(p)$.

Under the SDC, the Agent's reward is $Max\,\{0, T(\epsilon) - D - \psi(p)\}$. The Agent's expected profit is given by the following expression,

$$\Pi(D) = \mathbb{E}\big[Max\,\{0, T(\epsilon) - D - \psi(p)\}\big] \tag{12}$$

We find that the Agent sets the policy that the King would choose in the hypothetical first-best situation studied above, that is, $(p^*, t^*(\epsilon))$. We can state the following result.

[7] Gale and Hellwig (1985) define an SDC as follows: "a contract which requires a fixed repayment when the firm is solvent; and requires the firm to be declared bankrupt if this fixed payment cannot be met and allows the creditor to recoup as much of the debt as possible from the firm's assets."

Proposition 3. *Under an SDC, the private agent sets the policy* $(p^*, t^*(\epsilon))$ *that would be chosen by a perfectly informed profit-maximizing King.*

The proof is immediate since the private agent and the informed King share the same objective, up to a constant D, for each ϵ, that is, $T(\epsilon) - \psi(p)$.

4.1.2 The optimal lease

We first state a useful result.

Lemma 1. *If* D *is chosen in the interval,*

$$T^*(\epsilon_1) - \psi(p^*) < D < T^*(\epsilon_0) - \psi(p^*),$$

then there exists a function $\epsilon^*(D)$ *such that,*

$$T^*\left(\epsilon^*(D)\right) = D + \psi(p^*), \tag{13}$$

and $\epsilon_0 \leq \epsilon^*(D) \leq \epsilon_1$.

For a proof, see the appendix.

To simplify the notation, we denote $\epsilon^*(D) = \epsilon^*$. Remark that, differentiating ϵ^* with respect to D gives $\epsilon^*_D = 1/T^*_\epsilon(\epsilon) < 0$.

The Agent's expected surplus can be rewritten,

$$\Pi(D) = \int_{\epsilon_0}^{\epsilon^*(D)} T^*(\epsilon) f(\epsilon) d\epsilon - \left(D + \psi(p^*)\right) F\left(\epsilon^*(D)\right) \tag{14}$$

The King chooses D to maximize his expected surplus, denoted V^F (where F stands for "farmer"),

$$V^F(D) = DF\left(\epsilon^*(D)\right) + \int_{\epsilon^*(D)}^{\epsilon_1} \left(T^*(\epsilon) - \psi(p^*) - K\right) f(\epsilon) d\epsilon, \tag{15}$$

subject to the participation constraint, $\Pi(D) \geq \pi$, where π is the Agent's reservation profit level. Define $D_{\max}$ as the value of D such that

$$\epsilon^*(D_{\max}) = \epsilon_0.$$

We require $D \leq D_{\max}$. Requiring a D larger than D_{max} would be useless, since the Agent would always be unable to repay the lease. Moreover, if

$\pi < 0$, then we have $\Pi(D) \geq 0 > \pi$ for any $D \leq D_{\max}$. This shows that the participation constraint cannot be binding if $\pi < 0$.

We can state the following result.

Proposition 4. *Let $D^* = D^*(K)$ be the solution of the equation*

$$\frac{f(\epsilon^*(D^*))}{F(\epsilon^*(D^*))} = \frac{-T^*_\epsilon(\epsilon^*(D^*))}{K}, \tag{16}$$

(if it exists). Then, the participation constraint is slack if $0 < \pi < \Pi(D^*(K))$ and the optimal lease is $D^*(K)$. Moreover, if $T^*_{\epsilon\epsilon}(\epsilon^*(D^*)) < 0$, then,

$$\frac{\partial D^*}{\partial K} < 0.$$

Otherwise, define $D^* = D^*(\pi)$ as the solution of the equation $\Pi(D(\pi)) = \pi$. Then, if

$$K \leq -T^*_\epsilon(\epsilon^*(D^*))\frac{F(\epsilon^*(D^*))}{f(\epsilon^*(D^*))},$$

the participation constraint is binding and the optimal lease is $D^*(\pi)$, with the property

$$\frac{\partial D^*}{\partial \pi} < 0.$$

In any case, the optimal lease is such that $D^* < D_{max}$.

For a proof, see the appendix.

Proposition 4 shows that the solution has two regimes. In the first regime, determined by (16), the participation constraint is slack: this corresponds to relatively small values of the reservation profit π. In the second regime, the participation constraint $\pi = \Pi(D)$ is binding: this corresponds to small values of K. It is possible to prove the existence of the solution in the first regime, at the cost of some standard technical assumptions. Intuitively, a solution exists since the left hand side of (16) is increasing with D, while the right hand side is decreasing if $T^*_{\epsilon\epsilon} < 0$: the two curves intersect at point $D^*(K)$.

The boundary between the two regimes is described, in the (K, π) plane, by the function $\pi^*(K) \equiv \Pi(D^*(K))$. This function is monotonic since $\partial\Pi/\partial D < 0$ and

$$\frac{\partial \pi^*}{\partial K} = \frac{\partial \Pi}{\partial D}\frac{\partial D^*}{\partial K} > 0.$$

The participation constraint is binding for points (K, π) above the π^* curve.

4.2 Delegation to a public tax collector under a PAC

We now turn to the case of a civil servant. The public agent is in charge of collecting taxes and detecting tax evasion as above. There is a fixed cost C, incurred in setting up an accounting system, that provides a measure of the fiscal revenues T. The civil servant chooses the effort level p, based on a private observation of the random shock ϵ. The King observes neither ϵ nor the agent's effort p. To keep things comparable, we assume that the civil servant chooses[8] t and p, and incurs the effort cost $\psi(p)$. The King proposes a linear contract, with compensation specified as $\alpha + \beta T$, to the civil servant, where α and β are parameters, to be determined, and the King receives $T(\epsilon)$.

The civil servant's utility is defined as

$$U(\epsilon) = \alpha + \beta T(\epsilon) - \psi(p), \tag{17}$$

where, to simplify notation, we use the shorthand notation, $U(\epsilon)$ and $T(\epsilon)$, as above, for $U(t, p, \epsilon)$ and $T(t, p, \epsilon)$, respectively. The civil servant's reservation utility is normalized to $u_0 = 0$. We consider a participation constraint of the civil servant, expressed in terms of expected utility, that is, we require,

$$\mathbb{E}U(\epsilon) \geq 0. \tag{18}$$

We also assume that the civil servant's liability is limited. The limited liability constraint (hereafter LL) is expressed as follows, $U(\epsilon) \geq R$, *for all* ϵ, where R is a given constant. The LL constraint will play a role similar to that of risk aversion in the standard Principal-Agent model. If $R > 0$, it is clear that LL implies (18).

4.2.1 First-best solution

Suppose for a moment that the King observes p and ϵ. For each ϵ, the King can then set $p(\epsilon)$ and $t(\epsilon)$ to maximize the expected profit,

$$V^B = (1 - \beta)\mathbb{E}T(\epsilon) - \alpha - C, \tag{19}$$

subject to the civil servant's participation constraint only, ignoring LL. Since V^B is a decreasing function of α, the participation constraint will be binding,

[8] Tax levels are usually at the discretion of the central authority and/or voted by Parliaments. This shortcut will greatly simplify the analysis.

implying $\alpha = \mathbb{E}[\psi(p(\epsilon)) - \beta T(\epsilon)]$. Thus, the King's expected profit can be rewritten as

$$V^B = \mathbb{E}[T(\epsilon) - \psi(p(\epsilon))] - C.$$

It follows that the King will maximize $T(\epsilon) - \psi(p)$ for each value of ϵ. This yields the same solution as in the statement of Proposition 2 above. The first-best profit is obtained with p^* and $t^*(\epsilon)$ satisfying the equations: $s = \psi'(p^*)$ and $\eta(\tilde{x}^*|\epsilon) = 1$ for all ϵ. We can now state the following Proposition.

Proposition 5. *If we ignore the LL constraint, then the first-best solution* $(p^*, t^*(\epsilon))$ *can be implemented by means of a linear contract* (α^*, β^*) *such that,*

$$\alpha^* = \mathbb{E}[\psi(p^*) - T^*(\epsilon)], \tag{20}$$

$$\beta^* = 1 \tag{21}$$

For a proof, see the appendix.

If we ignore LL, the Agent is risk-neutral and can be made the residual claimant of the King's tax revenues by setting the piece rate $\beta^* = 1$ and the base salary α^* that just satisfies the participation constraint. The civil servant then exerts the optimal effort level p^* and chooses the tax level $t^*(\epsilon)$ that maximizes the King's profit.

4.2.2 Second-best solution

Under LL, the Principal-Agent solution yields a second-best profit optimum for the King. The civil servant privately observes ϵ, and chooses $p(\epsilon)$ and $t(\epsilon)$ so as to maximize $U(p, t, \epsilon) = \alpha + \beta T(p, t, \epsilon) - \psi(p)$, with respect to (p, t), for all ϵ. We can now state a useful Lemma.

Lemma 2. *Assume that* $b(\epsilon) = +\infty$ *and* $\eta(\theta - s|\epsilon) > 1$. *Then, the policy variables* $(p(\epsilon), t(\epsilon))$ maximize $U(p, t, \epsilon)$ *if and only if the following conditions are satisfied*

$$\beta s = \psi'(p) \tag{IC_1}$$

$$\eta(t - ps|\epsilon) = 1, \tag{IC_2}$$

where η *is defined by (7) above. If* $\beta > 0$, *the solution is unique and interior,* i.e., *the only* $(p(\epsilon), t(\epsilon))$ *belongs to* $(0, 1) \times (0, \theta)$. *If* $\beta = 0$, *then* $p(\epsilon) = 0$ *and* t *can be chosen to satisfy* IC_2.

For a proof, see the appendix.

In technical terms, Lemma 2 says that the first-order approach is justified: IC_1 and IC_2 are equivalent to the Agent's incentive constraints.

We can now show that LL will always be binding at a second-best optimal solution. We can state the following result.

Lemma 3. *The limited liability constraint is binding at point ϵ_1, at the second-best optimal solution. If $\alpha = \psi(p(\epsilon_1)) - \beta T(\epsilon_1) + R$, then, the LL constraint is satisfied for all $\epsilon \leq \epsilon_1$.*

For a proof, see the appendix.

The King's problem is now to choose β so as to maximize

$$V^B = \beta T(\epsilon_1) - \psi(p) + (1-\beta)\mathbb{E}\left[T(\epsilon)\right] - R - C \tag{22}$$

subject to

$$\beta s = \delta p \tag*{(IC_1)}$$

$$1 - G\left(x|\epsilon\right) - xg\left(x|\epsilon\right) = 0 \tag*{(IC_2)}$$

where $x = t(\epsilon) - ps$ and it is understood that $T(\epsilon)$ means $T(p, t(\epsilon), \epsilon)$.

The next result gives the parameters of the second-best optimal contract with the public agent.

Proposition 6. *Under the limited liability constraint LL, the second-best optimal contract, denoted $(\overline{\alpha}, \overline{\beta})$ is such that*

$$\overline{\alpha} = \psi(\overline{p}) - \overline{\beta T}(\epsilon_1) + R, \tag{23}$$

$$1 - \overline{\beta} = \frac{\delta}{s^2}\left[\mathbb{E}\overline{T}\left(\epsilon\right) - \overline{T}\left(\epsilon_1\right)\right] > 0, \tag{24}$$

where $(\overline{p}, \overline{t})$ satisfy IC_1 and IC_2 and $\overline{T}(\epsilon)$ is shorthand for $T\left(\overline{p}, \overline{t}(\epsilon), \epsilon\right)$.

In particular, (24) shows that we have $\overline{\beta} < 1$.

For a proof, see the appendix.

Under asymmetric information and LL, we find that the piece rate $\overline{\beta}$ is smaller than 1, which is a departure from the first-best level, and the base wage $\overline{\alpha}$ is simply adjusted to meet the LL constraint for any ϵ.

The next Proposition shows the effort and the tax rate distortions resulting from the asymmetric information between the King and the civil servant, under LL.

Proposition 7: *Under the contract* $(\overline{\alpha}, \overline{\beta})$, *the optimal effort and tax levels are distorted downwards, as follows,*

$$\overline{p} = \overline{\beta} p^* < p^*, \tag{25}$$
$$\overline{t}(\epsilon) = \frac{1}{h(\overline{x}|\epsilon)} + s\overline{p} < t^*(\epsilon). \tag{26}$$

For a proof, see the appendix.

Proposition 7 illustrates the trade-off between rent-extraction and efficiency in a Principal-Agent model. In order to reduce the informational rent of the Agent, the King sacrifices efficiency to a certain extent. With $\overline{\beta} < 1$, the power of incentives is too low and both the effort and the tax rate are distorted downwards with respect to the first best solution. The civil servant has no wealth and is protected against penalties that would be too severe by the limited liability constraint. However, for incentive purposes, the King rewards his agent by means of a percentage of the tax revenue, which leaves the agent with an ex ante rent. If we substitute $\overline{\alpha} = \psi(\overline{p}) + R - \overline{\beta} T(\epsilon_1)$ into $\mathbb{E}U(\epsilon)$, we obtain,

$$\mathbb{E}U(\epsilon) = R + \overline{\beta}\Big(\mathbb{E}\overline{T}(\epsilon) - \overline{T}(\epsilon_1)\Big) > R$$

The agent earns R, as required by the LL constraint, plus a rent

$$\overline{\beta}\left[\mathbb{E}\overline{T}(\epsilon) - \overline{T}(\epsilon_1)\right],$$

which is positive because $\overline{T}$ is a decreasing function of ϵ.

To sum up, under asymmetric information, when the King chooses to delegate the collection of taxes to an agent, in both the civil-servant and the tax-farmer cases, an informational rent has to be left to the agent. Under the SDC, the tax farmer is made the residual claimant for the total revenues, and thus enjoys a rent, but makes profit-maximizing choices. Under the public tax collection regime, a rent has to be paid in order to satisfy the agent's LL constraint, and in spite of a better measurement of performance, the agent's policy choices are inefficient. We need to study how the recourse to a civil servant can be more profitable than tax-farming, given that a costly accounting system, yielding observations of the tax revenue, is required in the public agent case.

4.3 The choice between the private and public agent

The choice between the SDC and the PAC, and the transition from tax-farming to public management depend on the value of parameters K, C and R. If C and R are very large, it is clear that the King prefers the SDC. We want to compare the expected profits under the two contracts. We first assume that the state verification cost K is such that $\epsilon_0 < \epsilon^*(D) < \epsilon_1$, and the solution $D^* = D^*(K)$ is given by (16), in the statement of Proposition 4 above.

To compare the two forms of contract, we need a point of comparison. It seems natural to assume that $\pi = u_0 = 0$, that is, the tax-farmer and the civil servant have the same reservation profit. In this case, we know that the SDC solution is not constrained by the participation constraint, since $\Pi(D) \geq 0 = \pi$. The King's expected surplus under the SDC is denoted V^F, given by (15) above. Under the PAC, the King's expected surplus is V^B, given by (22) above. The King prefers the tax-farmer to the civil servant if and only if $V^F \geq V^B$. We can state the following result.

Lemma 4. *$V^F \geq V^B$ if and only if,*

$$\begin{aligned} T^*(\epsilon^*)F(\epsilon^*) + \int_{\epsilon^*}^{\epsilon_1} T^*(\epsilon)f(\epsilon)d\epsilon - K\left[1 - F(\epsilon^*)\right] \\ \geq \mathbb{E}\overline{T}(\epsilon) + (1-\overline{\beta})^2\frac{s^2}{2\delta} - R - C, \end{aligned} \tag{27}$$

with $\epsilon^* = \epsilon^*(D^*(K))$.

For a proof, see the appendix.

If $C + R$ is large, tax-farming is more profitable. The question of interest is whether it would be possible to save on the private agent's rent by the recourse to a public agent. The following result states the existence of a threshold for the verification cost, triggering the shift from the tax-farmer to the civil servant.

Proposition 8. *If $0 \leq R+C < (1-\overline{\beta})^2 s^2/2\delta$, then, there exists a threshold $\widehat{K}$, with $V^F(\widehat{K}) = V^B$, such that if $K \leq \widehat{K}$, the tax farmer is a better deal. Otherwise, if $K > \widehat{K}$, the civil servant is the best strategy for the King. The threshold $\widehat{K}$ is an increasing function of $R + C$.*

For proof, see the appendix.

This proposition states that neither of the contracts dominates the other, but that there are regions where one is more valuable than the other, depending on the values of K, C and R. It is worthwhile noting that when the fixed costs associated with the civil servant increase, that is, when $R+C$ increases, the verification cost threshold $\widehat{K}$ that leaves the King indifferent between the two arrangements increases. This is in line with intuition, as it becomes more costly to contract with the civil servant. In the next section, we confront the results of this theoretical analysis with some historical facts on the shift from one form of fiscal organization to the other.

5 Does the model explain historical facts?

Formal models motivated by historical analyses can help understand the variation of fiscal regimes across time and states. Unfortunately, most of the time, they are not able to account for all the dimensions at once or fit the classification of taxes very well. Let us first distinguish different categories of taxes. They may be direct or indirect, defined on a personal basis or levied on consumption. Taxes might be imposed on output (the harvests or the craft goods) or on the inputs entering the production processes (such as fruit trees or land). They can be fixed, if imposed on individuals or land, or variable if levied on output, traded or consumed goods. The tax base can be more or less volatile: customs revenues have more variability than land taxes for example. Finally, the skills of the personnel required for tax collection matter: customs duties involve heavy surveillance of the borders, while personal taxes and property tax collection require agents capable of accurately assessing the value of the tax base.

As emphasized by Coşgel (2005), all of these features generate different kinds of costs, attached to the measurement of the tax base, to tax collection and to the monitoring of tax collectors. Accounting for these costs improves the understanding of the variation in the fiscal systems across time and space. For the sake of clarity, recall that we use the respective terms SDC for the tax-farming type of contract, and PAC for the mode of tax collection involving public agents. It should be remarked that the PAC covers different arrangements such as the *régie,* a public company run by public agents earning a fixed salary ($\beta = 0$). The case of a *régie intéressée,* run by a public agent earning a percentage of the revenue, is described by $0 < \beta < 1$, as in our model. Finally, $\beta = 1$ corresponds to a situation, where everything works as if the King had sold the tax collection activity to a private agent in exchange for a certain payment.

History reveals that SDCs have been observed almost everywhere. The right to collect taxes was granted in two main ways. It could be adjudicated by means of competitive bidding: the private agent with the highest bid getting the lease. This system was used by the Roman Republic (third century BC), the Abbasid Empire, and other Islamic states, by the Ottomans in their early modern period, in England, in France (from the sixteenth century to the end of the *Ancien Régime*) and in other European countries during the same period. The right to collect taxes could also be negotiated between the two parties, as was done by the Ottomans and the French Kings who used competitive bidding until 1661 and negotiation after 1661. The PAC is the other arrangement observed. It became dominant at the end of the eighteenth century. If officials appear late in early modern Europe, the Ottomans had recourse to salaried agents, and as had the Chinese long before. Hereafter, we provide some elements concerning the respective fiscal organization of the countries under study.

5.1 Institutional context in France, England and precolonial Morocco.

The early modern French fiscal system was incredibly complex. The country was divided into provinces, differing in their degree of autonomy with respect to central authority, the *pays d'Etat*[9], *pays d'élection*[10], and *pays d'imposition*[11]. The applicable fiscal regimes varied across regions. The direct taxes were collected as follows: a royal officer– the *intendant des finances* – was to collect a given amount in his jurisdiction, assisted by the *receveurs généraux des finances* (general receivers of finance) who were in charge of the accounting task of centralizing the receipts by other agents (collectors at

[9]There, the representatives of the different social groups negotiated the taxes with the King.

[10]In the *pays d'élection*, each jurisdiction (*généralité*) was asked to pay an amount decided by the central authority.

[11]No autonomy was granted to those territories that had been recently conquered (in the seventeenth and eighteenth centuries) .

the parish level). The indirect taxes[12] were farmed out.

The British tax system had comparable salient features. The indirect taxes such as customs were farmed out, while the direct taxes were collected by local agents. As documented by Kiser and Kane (2001), private agents were in charge of the customs collection, their organization became centralized before the English Revolution of 1688, and was very quickly characterized by a bureaucratic hierarchy. The direct taxation system was organized in the same way as in France.

The Moroccan Monarchy's history has been marked by successive waves of invasions (among others, the Phoenicians, Romans and Arabs). The tribal warfare that prevailed for a long period impeded the unification of the country. Very early, it was divided into two distinct parts. In the territory called *Bilad al Makhzan*, where the monarch's authority was recognized, the citizens consented to taxes. The dissident tribes formed the other part, the *Bilad al Siba*. The main taxes in use in precolonial Morocco were the *zakat*, the *ushur* and the *naiba*. The *zakat*, originally a voluntary contribution paid by the citizens for charity purposes, became compulsory and levied on accumulated wealth. The *ushur* (tithe) was a tax on the harvest levied on the peasants alone, at the rate of 10%. The *naiba*[13] was a property tax used, when needed, to fuel the *bait-el-mal* (treasury), to fund the military expenditures. Additional taxes fell on the commercial activities. Some of them were considered illicit by religious representatives. The monarchs had officials who were local governors – the *qaids* in the countryside and the *pashas* in urban areas[14] – responsible for tax collection. The *pashas* collected, for instance, a tax on shopkeepers. The *qaids* raised various taxes, such as the *zakat*, the *ushur* and the customs duties *(maqs)* as well as the *jiziya*, paid by the Jews in exchange for protection. The *qaids* had the power to grant exemptions. They also provided the monarch with military troops for the tax-collection expeditions. As judges, dealing with both civil and criminal

[12]The most important were imposed on salt consumption (*gabelle*); on various products, including wine and beer (sales taxes called *aides*); customs duties at the borders (*traites*); tolls at the gates of Paris (*entrées*); and different regalian rights (*domaines*), which are revenues from assets such as water, forests, and registration fees.

[13]Its rate was fixed by the sultan's representative. This implied a large variability of its revenues.

[14]The Saads reigned from 1559 to 1659 and the Alawis from 1666 until today. Ahmed al Mansur, from the Saads (1578-1603), a contemporary of Elizabeth I, contributed to the unification of Morocco. He imported elements of governance such as the Pashaliks and the Qaidship from the Ottoman Empire.

issues, they could impose both monetary fines and imprisonment. Important for our discussion is the SDC nature of their contract with the monarch. The *qaids* had to pay a fixed amount to the sultan, and they were allowed to use the taxes and fines as their source of income[15]. The *qaids* were assisted by court officials, the *sheikhs*, who were responsible for gathering the receipts from local collectors, called the *djarys*. Since neither the collectors nor the court officials had a salary, they would pay themselves, and in the absence of monitoring of their activities, this generated losses.

5.2 A common historical pattern from private to public tax collection

The case studies throughout history display a common shift from the SDC to the wage contract, observed in England, as well as Sweden, Russia, and other European countries in the eighteenth century (see Bonney 1995). Hereafter we will concentrate on France, England and Morocco, concerned with the same transition, though at different periods of their history.

In France, SDCs were first used when the crown needed resources, mainly to finance war expenditures, and the crown's ability to borrow was especially low. A second move occurred in the eighteenth century, towards direct tax collection using public agents who had wage contracts. England decided to abandon the SDC in the late seventeenth century and built a fiscal administration with employees under wage contracts. In Morocco, the shift towards the wage contract occurred at the end of the French Protectorate, in 1956.

Our model suggests different forces that help explain why some states chose a privatized tax collection by entrepreneurs by means of SDC while others opted for bureaucracies of officials hired under PACs. The driving forces that we have identified are the incentive properties of SDCs, leading to efficient tax-collection technology; the costs K borne by the ruler to verify the tax revenues; the effort-monitoring costs of tax collectors C; the amount R necessary to protect collectors against variability in tax revenues and, finally, the harmful distortions of SDCs.

[15] See Bidwell (1973).

5.3 Incentive properties of the SDC

The first motive for resorting to SDCs, common to all the countries that had relied on this arrangement, is the security of the flow of revenues it provided them with. Moreover, as will be shown below, it is particularly striking that both England and France exploited the incentive properties of SDCs. As stated in Proposition 3, the private collector chooses the outcomes that a perfectly informed profit-maximizing King would. This gave the Kings the opportunity to engage in a profound reshaping of tax collection arrangements towards PACs.

During the fifteenth and sixteenth centuries, the increasing need for resources led the French crown to multiply taxes that were farmed out. Before 1661, the allocation of the farmers on the kingdom's territory was not optimal. It was fragmented, since the collection of the entry and exit taxes could involve multiple farms. This ended in inefficiencies in tax collection and led to the King's poor knowledge of the actual revenues. The concentration process began during the sixteenth century during Henri IV's reign. It was launched by Sully, his prime minister, who wanted to reduce the overhead costs of collection (see Frémy (1911)). This concentration was pushed under Colbert, Louis XIV's finance minister and made possible because tax farmers were entrepreneurs, able to accumulate the capital necessary to operate large farms. As of 1668, there existed 3 farms, each of them responsible for collecting the tax on salt and customs duties; the sales taxes (*aides*); and the domain rights, respectively. Since salt was taxed at different rates in different provinces, fraud was frequent; similarly, the customs duties levied at entries and exits were exposed to fraud and smuggling. Fighting against fraud and smuggling took the same efforts, so the collection and monitoring costs could be diminished by merging the collection of the tax on salt with the customs duties (see Pion (1902)). The cost-minimization logic was pushed until all the indirect taxes (on salt, trade, customs duties, domain and tobacco) ended up by 1691 being collected by a single syndicate of entrepreneurs, the *Ferme Générale*. The same argument regarding the tax collection technology can be used to explain the English experience. The concentration process of indirect tax collection had been at work in England too, starting from 1604, when the Great Farm concentrated the main customs revenues, until the 1640 consolidation of the Petty Farms and the Great Farm (for more details see Kiser and Kane (2001)).

In both countries, the concentration of the collection of indirect taxes in the hands of a single firm eased the bureaucratization. SDCs created strong incentives for the tax-farmers. Since they were the residual claimants of all the efficiency gains, they introduced efficient tax collection procedures and monitoring schemes. This is mainly due to a better protection of their property rights. Although, in principle, the lease price D was fixed during the contract period, in practice, discretionary renegotiation by the French crown generated insecure property-rights for the tax-farmers. The latter were sometimes unable to pay the lease price, because the King refused to reimburse the loans they had granted (see Johnson (2007)). As suggested by Proposition 4 above, the tax-farmers' participation constraint may have been binding, with an effective lease price corresponding to the constrained regime. Moreover, the threat of a hold-up is a likely explanation for the tax farmers' reluctance to invest in tax collection technology, resulting in large operating costs, which is evidenced by the small revenues of French farmers during the first half of the seventeenth century. The consolidation as a monopoly secured the farmers' property rights. Once competitive auctions had been abandoned, the tax collection industry became concentrated and better organized, allowing tax-farmers to negotiate more favorable lease prices[16]. Between 1661 and 1695, the revenues increased significantly (see Johnson (2006)) and the farmers enjoyed generous rents, a situation corresponding to the slack participation constraint regime of the SDC. This was to the advantage of the King but at the expense of the taxpayers. As stated in Proposition 3, the private collector, faced with an SDC, is led to the choices that a perfectly informed profit-maximizing King would make: a policy, according to Proposition 2, that entails over-deterrence and over-taxation with respect to a benevolent ruler's policy.

On the tax farmers' property rights, see North and Weingast (1989), who emphasize the checks and balances provided by the British institutions. They explain how, after the Glorious Revolution, the Parliament and an independent Judiciary limited the crown's ability to renege on its commitments. This endowed the crown with credibility concerning debt repayment. The English monarchs could then have access to credit at lower interest rates and the state could assume the risk attached to tax collection. These political events created room for a publicly organized fiscal system, based on salaried public officials. The English monarchs prompted the shift to public administration

[16] Being the bankers of the King also helped strengthen their bargaining position. See Balla and Johnson 2009 on the political power acquired by these financiers.

to avoid being too dependent upon powerful private tax collectors, and exploited the organizational innovations of tax farmers to bureaucratize tax collection. France followed the same path a century later, but the driving forces were the same. In Morocco, the *qaids*' property rights were obviously very fragile, since the sultans' governance rested on arbitrariness and confiscation of whatever wealth might have been accumulated by local governors, the latter being would-be rivals. This could explain the poor performance of the tax farming system in precolonial Morocco.

5.4 State verification costs K

As compared to England, France's territory was vast, divided into different regions with a varying degree of autonomy, and its fiscal system was barely unified, with feudal and royal taxes overlapping. England's fiscal system was centralized from a very early date as compared to France. England had a more homogeneous fiscal system and developed communication and transportation facilities sooner than France (Johnson and Koyama (2014)). The difference in size[17] of the two countries and the unequal development of communication and transportation systems explain the difference in verification costs. These costs were more likely to be higher in a state with a vast territory than in a smaller one. The analysis of Kiser and Kane (2001) suggests that the state verification costs K began to fall after 1640 in England. In France, this decrease was possible around 1780, thanks to the emergence of road infrastructures, which helped the development of communications and transportation systems. The reason why farmers were preferred by French kings can be found in Proposition 8 above, which determines a threshold level $\widehat{K}$ for the state verification cost which leaves the ruler indifferent between the SDC and the PAC. This threshold depends upon C and R, the costs associated with monitoring the collector's effort, and the minimal amount the collector should be paid given his limited-liability constraint. As a consequence, as long as both C and R were large, the threshold $\widehat{K}$ was large. As R and C decreased, the threshold $\widehat{K}$ began to decrease, so the move away from SDCs took place. The question is whether such drops in C and R were observed.

[17]See Szostak (1991)

5.5 The collectors' effort-monitoring cost C

As emphasized by Kiser (1994), the states where wage contracts were in use had strong bureaucracies, which could be due to their lower cost of monitoring the tax collectors' effort. According to Matthews (1958), high effort monitoring costs (large values of C in our model) could explain that the recourse to salaried officials was not possible. The taxes that were farmed out (such as excises and customs duties) needed intensive bureaucratic management that were not yet available. Before the mid-seventeenth century, conditions were not favorable for direct collection, in either France, or England. Indeed, according to Priks (2005), in mid-seventeenth century England the transition from SDCs to direct collection by public agents is linked with the monitoring costs faced by the crown.

In England, extensive monitoring of officials took place to discourage them to pocket the tax revenues. Brewer (1988) reports that the extensive monitoring could be seen in the significant increase in the number of monitors in charge of controlling tax collectors.

In France, 1716 is a turning point in the building of an information system. Under the combined influence of Antoine Pâris, then *receveur général* of direct taxes in the Dauphiné region, and his influential brother Claude, who was *Secrétaire du Roi*, the double-entry book-keeping system of accounting was introduced. Extensive monitoring took place in the collection of direct taxes. The general receivers of finance were first required to keep a day-book with an exact record of all their receipts and expenditures, and to forward a copy every two weeks to the Ministry of Finance. The general receivers were also financial intermediaries (see Mousnier (1979)): they centralized the payments made by the citizens and were responsible for reporting all the received amounts for the different taxes such as *taille*, *capitation,* and, later on, the *vingtième*[18], to the central authorities. They acted as bankers for the King, since they could make advances by issuing bonds (*rescriptions*) secured by the future revenues to be collected under their scrutiny. They favored more regular payments. Thanks to their organization, the centralization of the direct tax revenues was made possible. However, the centralization and

[18]These were direct taxes. The *taille* was a tax on the income (*taille personnelle*) and on land (*taille réelle*). The *taille* did not fall on everyone: the nobility and clergy were exempted. The *capitation*, introduced in 1695, was a tax levied on each individual and its amount depended on the class to which they belonged. It was abolished in 1789. Finally, all individuals had to pay the *vingtième,* created in 1749, which amounted to paying 5% of their total income.

the subsequent bureaucratization[19] were thwarted by the independence of the general receivers. Since they had bought their office, this patrimonial dimension acted as a strong barrier against reform until the end of venal offices in 1791. While Necker had reformed the *Ferme Générale* in 1780, his attempt to centralize and control the general receivers, replacing them by one central treasury under direct state control, failed.

It should be noted that France's *indirect* tax collection system underwent several transitions, back and forth, between SDCs and PAC (*régie* contracts). The severe fiscal shortfalls, triggered by adverse economic conditions and the increases in war expenditures (see White 2004), forced the crown to replace the SDC by the *régie* arrangement. This happened from 1709 until 1714 and from 1721 until 1726, just after John Law's famous bankruptcy in 1720. The *régie* episodes offered an opportunity to improve the governmental monitoring of fiscal resources.

The requirement of keeping day-books was then extended in 1721 to the *Ferme Générale*, at that time being under the *régie* arrangement. This change improved the monitoring activity: the continuous updates and reporting to the Finance Minister helped the government obtain a more precise measure of the fiscal revenues T. A significant decrease in C can be interpreted as a consequence of the establishment of improved information, *i.e.*, through the accounting system. The increased precision with which tax revenues could be observed made the use of direct public collection more profitable. Indeed, in standard Principal-Agent theory, when the performance measure T becomes less volatile, the power of incentives β can be increased. The Agent's informational rents, and, hence, the agency cost, can be reduced.

As explained above, the *Fermiers Généraux* adopted a more efficient mode of tax collection. As time passed, during the eighteenth century, the *Ferme Générale* became highly centralized and hierarchical, with a monitoring body in each local district, in charge of auditing the agents' activity, allowing for economies of scale in the auditing system. It became an efficient bureaucratic tax-collection institution, providing between 40 and 50% of the tax revenues. White (2004) gives further details on this organization. The stakes were too high not to be monitored by the crown: around 1788, the government also improved its monitoring technology, with officials in charge of auditing the farmers' activities. A hierarchy of agency relationships was in operation. The French Revolution eventually put an end to the *Ferme*

[19]See Kiser and Kane (2001)on the bureaucratization processes in England and France

Générale as a tax collection system when it was replaced with state-run tax collection.

The case of Morocco is interesting, insofar as it illustrates how the shift from the *qaids* and *pashas* system to public fiscal administration was the result of debt repayment monitoring by foreign lenders to the Moroccan government. During the nineteenth century, the presence of foreigners, backed by military forces, increased in the country[20]. The rivalry of Spain, France, Great Britain, and to some extent, Germany, was somehow moderated due to the establishment of international conventions and unequal trade treaties[21]. Different elements contributed to weakening the kingdom even further. First, the large openness of the country made it suffer structural trade deficits. Second, it experienced a public finance crisis with huge fiscal shortfalls due to periods of drought, to the generous tax exemption regime and to the protection system[22]. The wars with Spain, in 1860 and in 1894 were followed by heavy reparations that the Moroccan government was unable to pay without borrowing. In 1861, Morocco was granted a British loan. The external debt went on growing with the 1904 and 1910 French loans. By the end of Aziz's reign in 1908, the French presence in Morocco was significant. The protectorate took place from 1912 to 1956.

The country was finally mortgaged to European interests. The share of the internal debt was likely to be very small because of the ban, for religious reasons, on usurious interest rates. As a consequence, the Moroccan debt was owned by Europeans, and especially by the French. The role of the Moroccan public debt crisis[23] as an origin of the French protectorate is well recognized[24].

The debt repayment problem generated very strong monitoring incentives for Spanish and British governments that had sent controllers to secure repayments, paving the way for the French colonial administration. The British administration played a significant role in persuading the sultan Mohamed to review the fiscal system characterized by very poor yields (because

[20]The 1844 defeat (against French troops) took place in Isly near the Algerian frontier, where the Moroccan troops came to fight with their neighbors.

[21]For instance, the treaty of Algeciras in 1906.

[22]Thanks to the British treaty in 1856, foreigners were allowed to avoid Moroccan justice and paying taxes; and the Madrid convention in 1880 allowed some Moroccans (chosen by the foreigners) to enjoy the same advantage.

[23]The structural trade deficits were partially compensated by loans. See Barbe (2016) on Morocco's debt.

[24]See I. Lapidus (2002)].

of exemptions and pocketing by revenue collectors). In 1901, a new tax on agriculture was introduced, the *tertib,* which replaced the *zakat*, *ushur* and *naiba*. This came with a new organization of tax collection: no exemption could be granted, the tax collectors were sworn in and paid, and a controlling body was created, a team of *amin* and *adouls*, to assess and register the tax base.

The French administration shared the same objectives as the British. Even before the protectorate, following the 1904 loan, French officials began to monitor the Moroccan agents collecting customs duties. The monitoring became even more stringent after the 1910 loan, with the creation of the *Administration du Contrôle de la Dette Publique Mahghzenienne*, which was allowed to seize all the customs duties to service the debt, and continued its activity during the protectorate. The protectorate introduced a more systematic collection of taxes, thanks to a better organized administration, resulting in smaller monitoring costs and thus allowing the central government to closely control the *qaids*.

5.6 Variance of the tax base

Following the classification of taxes, the variance of the tax base is also relevant in explaining why some arrangements are preferred to others. The late sixteenth and seventeenth centuries were characterized by instability for European states. Rulers had to deal with hostile environments: adverse economic conditions which depressed growth and gave rise to inflation; political instability and religious tensions, the consequences of which were likely to increase the variance in the tax bases. This might have made SDCs more profitable for the collection of taxes whose bases were likely to be impacted by the environment changes. For instance, the revenue from customs is likely to be more variable than that from land taxes. The variance in the customs revenues is very much dependent on the economic conditions affecting international trade, such as exchange rates and the supply and demand levels. Customs duties were more sensitive to fluctuations in prices than excise taxes that were perceived within local markets. This larger sensitivity translated into a larger variance in the customs revenues, which calls for the SDC, because more volatility means more risks to be borne by the agent responsible for tax collection. If a risk-averse agent or an agent protected by limited liability, as in our model, was in charge of tax collection, then he would have had to be protected against this increased volatility. Resorting to public agents would have cost much more in more volatile environments. Resorting to tax farmers had the advantage that they could bear the risks attached to random variations in revenues.

As a consequence, whenever some activities experienced a decrease in the variance of the tax base, a move towards PAC could take place. During the eighteenth century, the agricultural and industrial revolutions respectively reduced the risk of crop failures, leading to lower variances of expected yields, and contributed to reducing the risks affecting production. Greater integration of markets reduced the effect of sector-specific shocks on revenues, again diminishing the variance in the tax base. It became less costly to hire public agents since they were less exposed to risk and the agency costs could be reduced.

Finally, there were other factors that favored the choice of PAC by making the alternatives less attractive, among them the social costs of tax collection caused by the abusive behavior of collectors.

5.7 Harmful distortions

The SDCs have had perverse consequences, such as over-taxation (or unequal treatment of citizens) or violent means of collecting taxes. Proposition 2 shows that delegation to a private agent leads to distortions in the form of over-deterrence and over-taxation with respect to the benevolent ruler's choices. The fact that farms were originally fragmented worsened the distortions. Each type of tax was differently farmed out by districts and the farmers would subcontract with other farmers at high prices. This pushed the subcontractors to be tough in collecting taxes. The brutality with which the police officers employed by the Company, the *gabelous*[25], extorted the French people was, as mentioned by historians, one major reason for the fall of the tax farmers[26]. Tax farming became an unpopular institution. In 1760, Mirabeau urged its abandon, following the physiocrats' view that indirect taxes were depressing economic activity. Claims of injustice in tax collection were filed, leading to social costs of collection. The competent courts of law, the *cours des aides*, had to litigate an increasing number of conflicts concerning the collection of taxes. The *cours des aides* issued numerous *remontrances* (admonitions) to the tax farmers and invited the King to put

[25] The Company was given police powers to fight against fraud and tax evasion.

[26] 28 tax farmers (out of 40 who were judged) were finally executed on the 8th of May, 1794, among them Antoine Lavoisier, the father of modern Chemistry.

an end to the tax farming system due to its lack of fairness. The increasing number of *remontrances* – in 1768, 1771, 1775 for the most famous – reflects the increasing intolerance[27] to abuses (see Decroix (2011)). Though the tax farmers could collect with low private costs, the way they collected the taxes led to social costs that the authorities had underestimated. In 1780, the partition of the *Ferme Générale* was decided by Necker.

In Morocco, the SDC between the Sultan and the *qaids* provided strong incentives, but at the cost of distortions: in some areas, overzealous *qaids* were famous for being particularly severe when rendering justice, and others for voluntarily triggering troubles in order to fine people. This type of contract was still in use under the Protectorate, during which the French administration tolerated some abuses[28]. Because injustice in enforcing the law and unfairness in tax collection were a source of instability, potentially impeding the state unification process. When the state recovered its independence in 1956[29] the *qaids* eventually became civil servants and were granted wages. As is predicted by the Theory of Incomplete Contracts (see, *e.g.*, Hart, Schleifer and Vishny 1997), if replacing a misbehaving agent is difficult and if the way the public task is undertaken matters, it is better to deal with a public agent rather than with a private agent who pursues private benefits.

6 Conclusion

We proposed a simple model, in which a central government delegates both tax collection and law enforcement. In one situation, the tasks were delegated to a private profit-maximizing agent-called the tax farmer- under a Standard Debt Contract à la Townsend (1979). This type of arrangement is known to be optimal when the observation of income is costly. However, we showed that this arrangement leads the King and the tax farmer to share the surplus at the expense of the taxpayers, who suffer over-taxation and distortions through law enforcement activities. Tax collection by means of civil servants, in contrast, is modeled by means of a Principal-Agent model in which the

[27]Malesherbes, the first president of the *Cour des aides de Paris* was one of the most severe opponents. In 1768, the abuses denounced concerning the tax farming system were severely criticized by the *cours des aides*.

[28]Sultan Mulay Hassan I (1873-94) was the first to consider a fiscal reform promoting more equality. He died during a tax collection expedition.

[29]Dahir du 7 chaabane 1375 (Decree of March 20th, 1956), an official ruling that defined the *qaids*' status.

civil servant signs a linear contract with the King. We then studied the theoretical conditions leading to a transition from tax-farming to public fiscal administration. Historical evidence supports the existence of the two forms of contract suggested by our model.

We have seen that in France and in Great Britain the cost minimization process favored concentration among the tax farms. The strong incentive properties of SDC were exploited: the concentration process fostered the decrease in operating costs and allowed the birth of bureaucratization, which in turn helped the shift towards better monitored collectors, eventually hired under a wage contract. We have also seen, for instance, that the tax farmers, in France and the *qaids* and *pashas*, in Morocco, were responsible for the distortions (over-deterrence and over-taxation), explained by our model, that eventually became a serious concern for the monarchs. Finally, the improved efficiency of accounting and auditing systems, leading to better measurement of the tax revenues, eased the transition towards public administration, again as predicted by our model.

However, our model cannot encompass the full variety of fiscal systems across regions and across time, calling for the consideration of additional elements that would enrich the analysis. Modeling the existence of representative institutions would give a political economy view of the fiscal mechanism design problem. Introducing cooperation could also be interesting since even authoritarian rulers of premodern states benefited from the support – whether explicit or implicit – of some groups of society, often elite groups seeking to benefit from fiscal regimes.

7 Appendix

7.1 Proof of Proposition 1

This benevolent planner chooses $p(\epsilon)$ and $t(\epsilon)$ in order to maximize the welfare function,

$$W(\epsilon) = \theta - \int_0^{t-ps} xg(x|\epsilon)dx - psG(\tilde{x}|\epsilon) - t\left(1 - G(\tilde{x}|\epsilon)\right),$$

where $\tilde{x} = t - ps$, and subject to the following budget constraint,

$$psG(\tilde{x}|\epsilon) + t\left(1 - G(\tilde{x}|\epsilon)\right) = \psi(p)$$

To characterize the first-best optimum of the benevolent planner, let γ denote the Lagrange multiplier of the budget constraint. We drop the argument ϵ of p, t, and γ to simplify notation. Assuming an interior solution, p^o, t^o, γ^o satisfy the following first-order necessary conditions:

$$\gamma^o\left[p^o sG(\tilde{x}^o|\epsilon) + t^o\left(1 - G(\tilde{x}^o|\epsilon)\right) - \psi(p^o)\right] = 0 \tag{28}$$

$$-\tilde{x}^o g(\widetilde{x}^o|\epsilon) + \gamma^o\left[1 - G(\tilde{x}^o|\epsilon) - \tilde{x}^o g(\tilde{x}^o|\epsilon)\right] = 0 \tag{29}$$

$$\tilde{x}^o g(\tilde{x}^o|\epsilon)s + \gamma^o\left[sG(\tilde{x}^o|\epsilon) + \tilde{x}^o g(\tilde{x}^o|\epsilon)s\right] = (1 + \gamma^o)\,\psi'^o) \tag{30}$$

where $\tilde{x}^o = t^o - p^o s$. From the second condition, we derive,

$$\frac{\tilde{x}^o g(\tilde{x}^o|\epsilon)}{1 - G(\tilde{x}^o|\epsilon)} = \frac{\gamma^o}{1 + \gamma^o} \geq 0$$

since $\gamma^o \geq 0$. The solution $\tilde{x}^o$ solves the equation

$$\frac{\gamma^o}{1 + \gamma^o}\left(\frac{1}{\tilde{x}^o}\right) = \frac{g(\tilde{x}^o|\epsilon)}{1 - G(\tilde{x}^o|\epsilon)},$$

Hence, the solution always exists and is interior since the right-hand side of the above equation is positive and increasing by Assumption 1 and the left-hand side decreases from $+\infty$ to 0. In addition, we must have $\gamma^o > 0$. Using the second optimality condition, we can rewrite the third one as follows,

$$\psi'^o) = \frac{\gamma^o}{1 + \gamma^o}s = \frac{\widetilde{x}^o g(\tilde{x}^o|\epsilon)}{1 - G(\tilde{x}^o|\epsilon)}s = \eta(\tilde{x}^o|\epsilon)s. \tag{31}$$

■

7.2 Proof of Proposition 2

The profit-maximizer chooses $p(.)$ and $t(.)$ to maximize

$$V(\epsilon) = psG(\tilde{x}|\epsilon) + t(1 - G(\tilde{x}|\epsilon)) - \psi(p).$$

Assuming that the solution exists and is interior, *i.e.*, $\tilde{x}^* > 0$, the first-order necessary conditions for optimality can be written as follows,

$$s[G(\tilde{x}^*|\epsilon) + \tilde{x}^* g(\tilde{x}^*|\epsilon)] = \psi'^*) \tag{32}$$

$$1 - G(\tilde{x}^*|\epsilon) - \tilde{x}^* g(\tilde{x}^*|\epsilon) = 0 \tag{33}$$

The second condition can be rewritten as $\eta(\tilde{x}^*|\epsilon) = 1$. This implies that $\tilde{x}^* > 0$. From the first necessary condition we then immediately derive $s = \psi'^*)$. It follows that p^* is a constant, independent of ϵ. We now show that $p^* > p^o$ and $t^* > t^o$ for all ϵ. First, since $p^* = (\psi'^{-1}(s)$ and $p^o = (\psi'^{-1}(s\eta(\tilde{x}^o|\epsilon))$, since $\eta(\tilde{x}^o|\epsilon) < 1$, and since $(\psi'^{-1}$ is strictly increasing, we immediately obtain $p^* > p^o$ for all ϵ.

Next, from Proposition 1, we know that $\eta(\tilde{x}^o|\epsilon) < 1 = \eta(\tilde{x}^*|\epsilon)$. By Assumption 3, η must be increasing with respect to x, for all ϵ. It follows that $\tilde{x}^o < \tilde{x}^*$, or equivalently, $t^o - t^* < (p^o - p^*)s < 0$. ■

7.3 Proof of Lemma 1

We start with the statement and proof of a useful result, that will be used in subsequent proofs.

Lemma 0. The optimal tax revenue, denoted $T^*(\epsilon)$, is a decreasing function of ϵ, that is,

$$T^*_\epsilon = -\tilde{x}^* G_\epsilon(\tilde{x}^*|\epsilon) < 0.$$

In addition, we have,

$$T^*_{\epsilon\epsilon} = (\tilde{x}^*_\epsilon)^2 [2g + \tilde{x}^* g_{\tilde{x}}] - \tilde{x}^* G_{\epsilon\epsilon}.$$

Proof of Lemma 0.

Differentiating $T^*(\epsilon)$ with respect to ϵ, using the shorthand notation $G^* = G(\tilde{x}^*|\epsilon)$, and $g^* = g(\tilde{x}^*|\epsilon)$, we obtain,

$$T^*_\epsilon(\epsilon) = [1 - G^* - \tilde{x}^* g^*]t^*_\epsilon + [G^* + \tilde{x}^* g^*]p^*_\epsilon s - \tilde{x}^* G_\epsilon(\tilde{x}^*)$$

Since $\eta(\tilde{x}^*|\epsilon) = 1$ and $p^*_\epsilon = 0$, we find $T^*_\epsilon(\epsilon) = -\tilde{x}^* G_\epsilon(\tilde{x}^*)$. Moreover, given that $\widetilde{x}^* > 0$ and by Assumption 1, $G_\epsilon\left(\tilde{x}^*\right) > 0$, we finally find $T^*_\epsilon\left(\epsilon\right) < 0$.

Differentiation with respect to ϵ yields the following second-order derivative,

$$T^*_{\epsilon\epsilon} = -\tilde{x}^*_\epsilon \left[G_\epsilon + \tilde{x}^* g_\epsilon\right] - \tilde{x}^* G_{\epsilon\epsilon} \tag{34}$$

Besides, we know that for all ϵ,

$$G\left(\tilde{x}^*|\epsilon\right) + \tilde{x}^* g\left(\tilde{x}^*|\epsilon\right) = 1. \tag{35}$$

Differentiating this expression with respect to ϵ, we obtain,

$$G_\epsilon + \tilde{x}^* g_\epsilon = -\left[2g + \tilde{x}^* g_{\tilde{x}}\right] \tilde{x}^*_\epsilon. \tag{36}$$

Therefore (34) becomes,

$$T^*_{\epsilon\epsilon} = \left(\tilde{x}^*_\epsilon\right)^2 \left[2g + \tilde{x}^* g_{\tilde{x}}\right] - \tilde{x}^* G_{\epsilon\epsilon}. \tag{37}$$

Remark that, by Assumption 4, we have $2g + \tilde{x}^* g_{\tilde{x}} > 0$. Hence, if $G_{\epsilon\epsilon} > 0$ is large enough, we have $T^*_{\epsilon\epsilon} < 0$. ■

Proof of Lemma 1.

Since by Lemma 0, we have $T^*_\epsilon < 0$, then, for D such that

$$T^*(\epsilon_1) - \psi(p^*) < D < T^*(\epsilon_0) - \psi(p^*),$$

by the Intermediate Value Theorem, there exists a function $\epsilon^*(D)$ such that,

$$T^*\left(\epsilon^*(D)\right) = D + \psi(p^*). \tag{38}$$

■

7.4 Proof of Proposition 4

Let λ and ρ be the multipliers associated with the participation and $D \leq D_{max}$ constraints, respectively. We maximize V^F subject to these constraints. The Lagrangian for this problem can be written as $L(D, \lambda, \rho) = V^F(D) + \lambda\left[\Pi(D) - \pi\right] + \rho\left[D_{\max} - D\right]$. Taking the derivative with respect to D and simplifying the expression yields the following Kuhn-Tucker conditions for optimality,

$$
\begin{array}{cccc}
& (1-\lambda)\,F(\epsilon^*) + Kf(\epsilon^*)\epsilon^*_D = \rho, & & \\
\lambda \geq 0, & \Pi(D) \geq \pi & \text{and} & \lambda\,(\Pi(D)-\pi) = 0, \\
\rho \geq 0, & D_{\max} \geq D & \text{and} & \rho\,(D_{\max}-D) = 0.
\end{array}
$$

These conditions define two regimes for the solution: a constrained and an unconstrained regime. The selection of a solution form depends on parameters π and K. Remark first, that $D = D_{\max}$ is not optimal. To see this, suppose the King sets $D = D_{\max}$, then $\epsilon(D_{\max}) = \epsilon_0$, $F(\epsilon_0) = 0$, and the first-order condition above boils down to,

$$\rho = Kf(\epsilon_0)\epsilon^*_D < 0,$$

since $\epsilon^*_D = 1/T^*_\epsilon < 0$. We have found a contradiction since we must have $\rho \geq 0$. We conclude that $D^* < D_{max}$.

Given that $\rho = 0$, the first-order condition above yields,

$$(1-\lambda)\,F(\epsilon^*) + \frac{Kf(\epsilon^*)}{T^*_\epsilon\,(\epsilon^*)} = 0.$$

The participation constraint may or may not be binding at the optimal solution, depending on the values of parameters π and K. In the unconstrained regime, the participation constraint is slack. This implies $\lambda = 0$ and the solution, denoted $D^* = D^*(K)$ is determined by the equation,

$$T^*_\epsilon\,(\epsilon^*(D)) + \frac{Kf(\epsilon^*\,(D))}{F(\epsilon^*\,(D))} = 0. \tag{39}$$

It follows that the participation constraint is slack if π satisfies $0 < \pi < \Pi(D^*(K))$.

Assuming that the solution exists, we can apply the Implicit Function Theorem to determine the sign of $\partial D^*/\partial K$. To simplify notation, define $\phi(\epsilon) = f(\epsilon)/F(\epsilon)$. The equation under study is $K\phi + T^*_\epsilon = 0$. Differentiating with respect to D yields,

$$\frac{\partial D^*}{\partial K} = \frac{-\phi}{(K\phi_\epsilon + T^*_{\epsilon\epsilon})\epsilon^*_D}. \tag{40}$$

Given our assumptions above, we have $\epsilon_D^* < 0$, $\phi_\epsilon < 0$ and $T_{\epsilon\epsilon}^* < 0$. Hence, $\partial D^*/\partial K < 0$.

In the constrained regime, the participation constraint is binding, we have $\Pi(D) = \pi$. It is easy to check that

$$\frac{\partial \Pi}{\partial D} = -F(\epsilon^*(D))$$

The participation constraint determines a solution denoted $D^* = D^*(\pi) = \Pi^{-1}(\pi)$. Then, the Kuhn and Tucker conditions imply that we must have,

$$0 \leq \lambda = 1 + \frac{f(\epsilon^*)}{F(\epsilon^*)} \frac{K}{T_\epsilon^*(\epsilon^*)}.$$

where $\epsilon^* = \epsilon^*(D^*(\pi))$. For a given π, there exists an interval of values of K, namely,

$$0 \leq K \leq -\frac{F(\epsilon^*)}{f(\epsilon^*)} T_\epsilon^*(\epsilon^*),$$

for which $0 \leq \lambda \leq 1$.

By the Inverse Function Theorem, we have

$$\frac{\partial D^*}{\partial \pi} = \left(\frac{\partial \Pi}{\partial D}\right)^{-1} < 0.$$

■

7.5 Proof of Proposition 5

Under the contract (α^*, β^*), the agent will choose $(p(.), t(.))$ to maximize $\alpha^* + T(\epsilon) - \psi(p(\epsilon))$. The civil servant will maximize $T(\epsilon) - \psi(p)$ and by Proposition 2, the solution is given by $p(\epsilon) = p^*$ and $t(\epsilon) = t^*(\epsilon)$. The agent's expected utility is therefore $\mathbb{E}U(\epsilon) = \alpha^* + \mathbb{E}T^*(\epsilon) - \psi(p^*) = 0$, so that the participation constraint is satisfied. ■

7.6 Proof of Lemma 2

The utility $U(p, t, \epsilon)$ is continuous with respect to (p, t), for all ϵ, over the compact set $A = [0, 1] \times [0, \theta]$. It follows that the problem $\max_{(p,t)\in A} U(p, t, \epsilon)$ has at least one solution $(p(\epsilon), t(\epsilon))$ for each ϵ. It is not difficult to check that this solution cannot belong to the boundary of A. We will now prove that for all ϵ, any interior local extremum of $U(p,t,\epsilon) = \alpha + \beta T(p,t,\epsilon) - \psi(p)$ with respect to (p,t) is a strict local maximum. The gradient of U is given by

$$\begin{aligned} U_p &= \beta s(G + \tilde{x}g) - \psi'(p), \\ U_t &= \beta(1 - G - \tilde{x}g). \end{aligned}$$

Denoting $I = 2g + \tilde{x}g_x$, simple computations show that the Hessian matrix, $H = \partial^2 U$, can be expressed as follows,

$$H = \begin{pmatrix} -\beta s^2 I - \delta & \beta s I \\ \beta s I & -\beta I \end{pmatrix}$$

Suppose first that $\beta > 0$ and consider a critical point (p,t) of U. We have $U_p = 0$ and $U_t = 0$. By Assumption 4, $U_t = 0$ implies $I > 0$. Hence, the determinant $det\,(H) = \delta\beta I > 0$ and the trace $Tr\,(H) = -\beta s^2 I - \delta - \beta I < 0$ have he required signs. The Hessian H is negative definite, indicating that the critical point (p,t) is a strict local maximum. If $\beta = 0$, we have $det(H) = 0$ and $Tr(H) = -\delta < 0$ for any (p,t). Hence, U is concave. In any case, IC_1 and IC_2 characterize a global maximum of U.

We finally show that the solution of $U_t = 0$, $U_p = 0$ is unique for $\beta > 0$. Condition IC_2 can be rewritten $h(x|\epsilon) = 1/x$. Under Assumption 3, it is easy to see that the latter equation has a unique solution $\tilde{x}(\epsilon)$. Equation IC_1 then determines a unique value of $p(\epsilon) = \beta s/\delta$. The value of p being pinned down, we obtain a unique $t(\epsilon) = \beta s/\delta + \tilde{x}(\epsilon)$. If $\beta = 0$, IC_1 yields $p = 0$, IC_2 is trivially satisfied, but we can in particular pick $t(\epsilon)$ such that $\eta(t(\epsilon)|\epsilon) = 1$ as a solution. ∎

7.7 Proof of Lemma 3

To prove the result, it is enough to show that $U(\epsilon)$ is a non-increasing function of ϵ. To see this, let $U(\epsilon) = \alpha + \beta T(p(\epsilon), t(\epsilon), \epsilon) - \psi(p(\epsilon))$ with $p\,(\epsilon)$ and $t\,(\epsilon)$ satisfying IC_1 and IC_2. We can therefore write the following,

$$U_\epsilon = \beta\left[T_p p_\epsilon + T_t t_\epsilon + T_\epsilon\right] - \psi'(p)p_\epsilon$$

Since we have $T_t = 0$, $\beta T_p - \psi' = 0$, and $T_\epsilon = -\tilde{x}G_\epsilon\,(\tilde{x}|\epsilon)$, the above expression boils down to,

$$U_\epsilon = \beta T_\epsilon = -\beta\tilde{x}G_\epsilon\,(\tilde{x}|\epsilon) \leq 0.$$

∎

7.8 Proof of Proposition 6

By Lemma 3, we can substitute LL into the objective. Hence, we maximize the following expression of V^B with respect to β only,

$$V^B = \beta\overline{T}\left(p, t(\epsilon_1), \epsilon_1\right) - \psi(p) + (1-\beta)\mathbb{E}\overline{T}\left(p, t(\epsilon), \epsilon\right) - C - R.$$

subject to $(p, t(\epsilon))$ satisfying IC_1 and IC_2. If $\beta > 0$, we can solve IC_1 and IC_2 for the unique solution $(\bar{p}(\beta), \bar{t}(\beta, \epsilon))$. By the Implicit Function Theorem, this solution is differentiable. Define

$$\overline{\overline{x}} = \bar{t}(\beta, \epsilon) - \bar{p}(\beta)s.$$

We now write the first-order necessary conditions, for an interior solution $0 < \overline{\beta} < 1$, as follows,

$$\overline{T}(\epsilon_1) + \left[\beta\overline{T}_p(\epsilon_1) - \psi'\right]\overline{p}_\beta + \beta\overline{T}_t\left(\epsilon_1\right)\bar{t}_\beta - \mathbb{E}\overline{T}(\epsilon) + (1-\beta)\mathbb{E}\left[\overline{T}_p\overline{p}_\beta + \overline{T}_t\bar{t}_\beta\right] = 0.$$

Under IC_1 and IC_2, we have $\beta\overline{T}_p(\epsilon_1) = \psi'$ and $\overline{T}_t = 0$. As a consequence, the first-order condition boils down to,

$$\overline{T}(\epsilon_1) - \mathbb{E}\overline{T}(\epsilon) + (1-\overline{\beta})\mathbb{E}\overline{T}_p\overline{p}_\beta = 0.$$

But, under IC_1 and IC_2, we have $\overline{T}_p = s$ and $\bar{p}(\beta) = \beta s/\delta$. Using these properties, we finally obtain,

$$1 - \overline{\beta} = \frac{\delta}{s^2}\Big(\mathbb{E}\overline{T}\left(\epsilon\right) - \overline{T}\left(\epsilon_1\right)\Big).$$

To establish that $\overline{\beta} < 1$, using a similar line of reasoning, we can write,

$$\frac{d\overline{T}(\epsilon)}{d\epsilon} = \overline{T}_p\overline{p}_\epsilon + \overline{T}_t\bar{t}_\epsilon + \overline{T}_\epsilon = \overline{T}_\epsilon = -\overline{\overline{x}}G_\epsilon < 0.$$

This result implies $\mathbb{E}\overline{T}\left(\epsilon\right) > \overline{T}\left(\epsilon_1\right)$. Hence $\overline{\beta} < 1$. ■

7.9 Proof of Proposition 7

Firstly, by IC_1, it is immediate that,

$$\overline{p} = \overline{\beta}s/\delta < p^* = s/\delta$$

since by Proposition 6, we have $\overline{\beta} < 1$. Secondly, by IC_2 we have $\tilde{x}^*(\epsilon) = \overline{\overline{x}}(\epsilon)$, because both functions are the unique solutions of the equation $h(x|\epsilon) = 1/x$. Therefore, for all ϵ, we have $\bar{t} - s\overline{p} = t^* - sp^*$ and we conclude that for all ϵ,

$$0 < s(p^* - \overline{p}) = t^* - \bar{t}.$$

■

7.10 Proof of Lemma 4

We denote, $\epsilon^*(D) = \epsilon^*$. Using (15) and the fact that, by definition of ϵ^*, $D + \psi(p^*) = T^*(\epsilon^*)$, we obtain,

$$V^F = T^*(\epsilon^*)F(\epsilon^*) - \psi(p^*) - K\left[1 - F(\epsilon^*)\right] + \int_{\epsilon^*}^{\epsilon_1} T^*(\epsilon)f(\epsilon)d\epsilon \tag{41}$$

Under the contract with a public agent, the King's expected surplus, given by (22) above, can be expressed as follows,

$$V^B = \overline{T}(\epsilon_1) - \psi(\overline{p}) + (1-\overline{\beta})\left(\mathbb{E}\overline{T}(\epsilon) - \overline{T}(\epsilon_1)\right) - R - C. \tag{42}$$

Now we can substitute $p^* = s/\delta$ and $\overline{p} = (\overline{\beta}s)/\delta$ in the inequality $V^F \geq V^B$. In addition, in the statement of Proposition 6, expression (24) yields,

$$(1-\overline{\beta})(\mathbb{E}\overline{T}(\epsilon) - \overline{T}(\epsilon_1)) = (1-\overline{\beta})^2 s^2/\delta.$$

Using (24) again, we also substitute

$$\overline{T}(\epsilon_1) = \mathbb{E}\overline{T}(\epsilon) - (1-\overline{\beta})\frac{s^2}{\delta},$$

to eliminate $\overline{T}(\epsilon_1)$. Putting these elements together in $V^F \geq V^B$ and simplifying yields (27). ■

7.11 Proof of Proposition 8

Given the result of Lemma 4, $V^F \geq V^B$ is equivalent to,

$$\begin{aligned} T^*(\epsilon^*)F(\epsilon^*) + \int_{\epsilon^*}^{\epsilon_1} T^*(\epsilon)f(\epsilon)d\epsilon - K\left[1 - F(\epsilon^*)\right] - \mathbb{E}\overline{T}(\epsilon) \\ \geq (1-\overline{\beta})^2\frac{s^2}{2\delta} - R - C, \end{aligned} \tag{43}$$

with $\epsilon^* = \epsilon^*(D^*(K))$. The RHS of the inequality is independent of K. Let the LHS of (43) be denoted by $\Phi(K)$. This function is decreasing with K. We have,

$$\frac{\partial\Phi(K)}{\partial K} = \left[T^*_\epsilon + K\frac{f(\epsilon^*)}{F(\epsilon^*)}\right] F(\epsilon^*)\,\epsilon^*_K - (1 - F(\epsilon^*))$$

Using (16), this leads to

$$\frac{\partial \Phi(K)}{\partial K} = -\left(1 - F\left(\epsilon^*\right)\right) < 0.$$

Next, we obtain,

$$\Phi(0) = T^*(\epsilon_0)F\left(\epsilon_0\right) + \int_{\epsilon_0}^{\epsilon_1} T^*(\epsilon)f(\epsilon)d\epsilon - \mathbb{E}\overline{T}(\epsilon) = \mathbb{E}T^*(\epsilon) - \mathbb{E}\overline{T}(\epsilon).$$

From the proof of Proposition 7, we know that $t^* - sp^* = \bar{t} - s\overline{p}$. This allows us to write,

$$T^*(\epsilon) - \overline{T}(\epsilon) = t^* - \bar{t} = s(p^* - \overline{p}), \tag{44}$$

and finally,

$$\Phi(0) = s(p^* - \overline{p}) = (1 - \overline{\beta})s^2/\delta.$$

Moreover,

$$\Phi(\overline{K}) = T^*(\epsilon_1)F\left(\epsilon_1\right) - \mathbb{E}\overline{T}(\epsilon) = T^*(\epsilon_1) - \mathbb{E}\overline{T}(\epsilon),$$

where

$$\overline{K} = -T^*_\epsilon \frac{F\left(\epsilon_1\right)}{f\left(\epsilon_1\right)}.$$

Using the result of Proposition 6, we can write,

$$\mathbb{E}\overline{T} = \overline{T}(\epsilon_1) + (1 - \overline{\beta})s^2/\delta.$$

We substitute the above expression into $\Phi(\overline{K})$. Using (44), we then find,

$$\Phi(\overline{K}) = T^*(\epsilon_1) - \overline{T}(\epsilon_1) - (1 - \overline{\beta})s^2/\delta = 0.$$

If the costs $R + C$ are such that

$$\Phi(\overline{K}) < (1 - \overline{\beta})^2 s^2/2\delta - (R + C) < \Phi(0),$$

then, there exists a $\widehat{K}$ such that $\Phi(\widehat{K}) = \left(1 - \overline{\beta}\right)^2 s^2/2\delta - (R + C)$. As a consequence, $V^F \geq V^B$ for all $K \leq \widehat{K}$.

Remark that, if $R + C = 0$, we have

$$\Phi(\widehat{K}) = (1 - \overline{\beta})^2 \frac{s^2}{2\delta} < (1 - \overline{\beta}) \frac{s^2}{\delta} = \Phi(0),$$

since $0 < 1 - \overline{\beta} < 1$. As a consequence, $\widehat{K} > 0$.

The recourse to a civil servant would dominate the SDC for all values of K if $(1 - \overline{\beta})^2 s^2/2\delta - (R + C) \geq \Phi(0)$, a case that we can rule out because it would imply $R + C < 0$.

Similarly, with $(R + C) = (1 - \overline{\beta})^2 s^2/2\delta$, then, $\Phi(\widehat{K}) = 0$ or $\widehat{K} = \overline{K}$. In this case, the SDC would always dominate.

As a conclusion, if

$$0 \leq R + C < (1 - \overline{\beta})^2 s^2/2\delta,$$

the threshold is such that $0 < \widehat{K} < \overline{K}$.

Finally, from $\Phi(\widehat{K}) = (1 - \overline{\beta})^2 s^2/2\delta - (R + C)$, we derive

$$\partial \widehat{K}/\partial(R + C) = -1/\Phi_K > 0.$$

■

8 References

Baker G. P. (1992). Incentive Contracts and Performance Measurement. *The Journal of Political Economy*, 100 (3), 598–614.

Baker G. P., Gibbons, R., Murphy, K. J. (1994). Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics* 109 (4), 1125–1156.

Balla E. and Noel Johnson, (2009). Fiscal Crisis and Institutional Change in the Ottoman Empire and France. *The Journal of Economic History*, 69 (3), 809–845.

Barbe, A. (2016). Public debt and European expansionism in Morocco From 1860 to 1956, mimeo, Paris School of Economics.

Becker, G. (1968). Crime and Punishment: An Economic Approach. *The Journal of Political Economy* 76 (2), 169–217.

Becker G., Stigler G. J. (1974). Law Enforcement, Malfeasance, and Compensation of Enforcers. *Journal of Legal Studies* 3, 1–18.

Bidwell R. (1973). *Morocco under Colonial Rule, French Administration of Tribal Areas, 1912-1956.* London, F. Cass.

Bonney R. (eds) (1995). *Economic Systems and State Finance. The Origins of the Modern State in Europe 13th to 18th Centuries.* Oxford University Press: Oxford.

Brewer J., (1988). *The Sinews of Power, War, Money and the English State 1688-1783.* Harvard University Press, Cambridge Mass.

Coşgel M. M. ,(2005), Efficiency and continuity in public finance: the Ottoman system of taxation, *International Journal of Middle East Studies*, 37 (4), 567–586

Coşgel M. M. , Haggay E., Miceli T. J. (2011). Private law enforcement, fine sharing, and tax collection: Theory and historical evidence. *Journal of Economic Behavior and Organization*, 80 (3), 546–552.

Decroix, A. (2011), Les Parlements, la réforme fiscale et l'opinion publique dans les dernières décennies de l'Ancien Régime,
Parlement[s], Revue d'histoire politique, 15 (1), 92-104.

Dharmapala Dhammika, Garoupa, Nuno, McAdams, Richard. H., (2016). Punitive police? Agency costs, law enforcement, and criminal procedure. *The Journal of Legal Studies*, 45 (1), 105–141.

Donahue, J. D. (1989). *Privatization Decision: Public Ends, Private Means.* New York: Basic Books.

Fremy E., (1911), Premières tentatives de centralisation des impôts indirects (1584-1614). In: Bibliothèque de l'école des chartes, tome 72. pp. 600-625.

Friedman, D. (1984). Efficient institutions for the private enforcement of law. *Journal of Legal Studies* 13 (2), 379–397.

Friedman D. (1999). Why not hang them all: the virtues of inefficient punishment. *Journal of Political Economy*, 107(S6), S259–S269.

Gale D., Hellwig, M. (1985). Incentive-Compatible Debt Contracts: The One-Period Problem. *The Review of Economic Studies* 52 (4), 647–663.

Garoupa, N. (1997). The Theory of Optimal Law Enforcement. *Journal of Economic Surveys*, 11 (3), 267–295.

Garoupa, Nuno, Klerman, Daniel, M., (2002). Optimal law enforcement with a rent-seeking government. *American Law and Economics Review* 4 (1), 116–140.

Grossman, S. J., Hart, O. (1983). An Analysis of the Principal-Agent Problem. *Econometrica* 51 (1), 7–45.

Hart, O., Shleifer, A. and Vishny, R. W. (1997). The proper scope of government: theory and an application to prisons. *Quarterly Journal of Economics* 112 (4), 1126–61.

Johnson, N.D. (2007). Banking on the King: The Evolution of the Royal Revenue Farms in Old Regime France. *The Journal of Economic History*, 66 (4), 963–991.

Johnson N. D., Koyama M. (2014). Tax farming and the origins of state capacity in England and France. *Explorations in Economic History* 51, 1–20.

Kiser E. (1994). Markets and hierarchies in early modern tax systems: a principal-agent analysis. *Politics and Society* 22 (3), 284–315.

Kiser E. and Kane J., (2001), Revolution and State Structure: The Bureaucratization of Tax Administration in Early Modern England and France, *American Journal of Sociology*, 107 (1), 183–223.

Landes W. M., Posner R. (1975). The private enforcement of law. *Journal of Legal Studies* 4 (1), 1–46.

Lapidus, I. (2002). A history of Islamic societies, Cambridge University Press, Cambridge (second edition).

Mathias, P, O'Brien, P.K. (1976), Taxation in Britain and France, 1715-1810: a comparison of the social and economic incidence of taxes collected for the central governments, *Journal of European economic history, 5: 601-50.*

O'Brien, P. K. (1988), The Political Economy of British Taxation, 1660-1815, The Economic History Review, 41 (1): 1–32.

Matthews, G. (1958). *The Royal General Farms in Eighteenth-Century France*. New York: Columbia University Press.

Mousnier, R. (1974), *Les institutions de la France sous la monarchie absolue* 1598-1789. Presses Universitaires de France, Collection "Histoire des institutions", Paris.

North, D. C., Weingast, B. R. (1989). Constitutions and Commitment: Evolution of Institutions Governing Public Choice in 17th Century England. *Journal of Economic History* 49: 803-832.

Pion, J.F. (1902), *La Ferme Générale des droits et domaines du Roi depuis sa création jusqu'à la fin de l'Ancien Régime*, V. Girard & E. Brière (eds), Paris

Polinsky, A. M., Shavell, S. (2007). The Theory of Public Enforcement of Law, *Handbook of Law and Economics*. In: Polinsky, A.M.,Shavell, S. (Eds), Handbook of Law and Economics, Vol. 1. North-Holland, Amsterdam.

Priks M. (2005). "Optimal Rent Extraction in Pre-Industrial England and France – Default Risk and Monitoring Costs," CESifo Working Paper Series 1464, CESifo Group Munich.

Stasavage, D. (2015). Why did public debt originate in Europe? In A. Monson & W. Scheidel (Eds.), Fiscal Regimes and the Political Economy of Premodern States (pp. 523-534). Cambridge: Cambridge University Press.

Szostak, Rick. 1991. *The Role of Transportation in the Industrial Revolution: A Comparison of England and France.* London: McGill-Queen's University Press.

Tilly, C., ed. (1975). *The Formation of National States in Western Europe.* Princeton, N.J.: Princeton University.

Tirole, J. (2006). *The Theory of Corporate Finance.* Princeton University Press, Princeton, New Jersey.

Toma E. F., Toma M. (1992). Tax Collection with Agency Costs: Private Contracting or Government Bureaucrats. *Economica* 59, 107–120.

Townsend R. (1979). Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory* 21, 417–425.

White E. N. (2004). From Privatized to Government-administered Tax Collection: Tax Farming in Eighteenth-century France, *Economic History Review* 57 (4), 636–663