Identification and Estimation of Incentive Contracts under Asymmetric Information: An Application to the French Water Sector [*]

Christian Bontemps[†]   David Martimort [‡]

Preliminary and Incomplete version

Abstract

We develop a Principal-Agent model to represent management contracting for public-service delivery. A firm (the Agent) has private knowledge of its marginal cost of production. The local public authority (the Principal) cares both on the consumers net surplus from consuming the services but also on the (weighted) firm's profit. Contractual negotiation is modeled as the choice by the privately informed firm within a menu of options determining both the unit-price charged to consumers and fixed fee. Our theoretical model characterizes optimal contracting in this environment. We then explicitly study the nonparametric identification of the model and rule a semi-parametric estimation on a dataset coming from the 2001 wave of a survey from the French environment Institute (IFEN, Institut Français de l'Environnement).

Keywords: Principal-Agent, optimal contracts, structural model, nonparametric estimation, Instrumental Variable Quantile Regression.

JEL codes: C12, C15, D82.

[†]Toulouse School of Economics (IDEI, GREMAQ). christian.bontemps@tse-fr.eu
[‡]Paris School of Economics-EHESS. david.martimort@parisschoolofeconomics.eu

1 Introduction

The increasing participation of the private sector in public-service delivery is often motivated by the need to expand access to services, increase or update existing delivery networks, and operate public utilities more efficiently. Many services provided by public utilities being associated with health, environmental or household income considerations (gas, electricity, water, transportation), there is a broad consensus on the need for public regulation of these utilities. In particular, industries such as water, gas and electricity usually meet the conditions for a local natural monopoly (large fixed costs and constant or declining marginal cost), so that protecting consumers from large price increases is often advocated as the main reason behind public regulation of utilities in these industries.

Private-sector participation in public utilities may take very different forms: private ownership of networks and facilities, centralized or local regulation by a public or independent authority, or contracting-out utility operation to private companies. In the latter case, typical arrangements are lease or concession contracts which can be renegotiated over time between a local community and a private company in charge of operating the public utility. Although contracting-out seems an interesting way of promoting public-private partnerships for public utilities because it combines flexibility with legal commitment, it can deteriorate consumer welfare if the arrangement concerning utility pricing is not carefully specified. A major reason behind the difficulty to design an optimal pricing rule for the utility is the fact that common information on the operator's ability to manage the utility efficiently is rare. For example, the operator's technical know-how and expertise may not correspond exactly to the actual state of the facilities (based on past maintenance). Such a situation of asymmetric information on the operator's efficiency in a contract-based relationship has been studied extensively in the literature on incentives and regulation (Salanié 2005, Laffont and Martimort 2002, Laffont and Tirole 1994).

In the standard theory of contracts, agents (operators in our case) are indexed by a private-information parameter which ultimately determines their actions, within a contract-based relationship with a "principal" (the local community). Whether this parameter denotes an unobservable action (moral hazard) or an unknown characteristic of the agent (adverse selection), the principal is assumed to have prior information used to design an optimal contract (in most cases, maximizing social welfare). The range of industries in which production, marketing and regulation activities are subject to contract-based relationships between economic agents is sufficiently large to guarantee an increasing number of empirical applications for such a theory.

2

The first approach in the literature on econometric estimation of delegated management models with asymmetric information was considering reduced-form models, or structural versions with restrictive parametric restrictions on the technology and the distribution of the private parameter. Examples of this first generation of models with asymmetric information include Wolak (1994), Ivaldi and Martimort (1994), Thomas (1995), Lavergne and Thomas (2005), Gagnepain and Ivaldi (2002), Brocas, Chan and Perrigne (2006). See Chiappori and Salanié (2000) for references on reduced-form estimation of models with asymmetric information, and Lavergne and Thomas (2005) for a survey on structural and reduced-form models. Most empirical applications confirm the fact that neglecting asymmetric information in the estimation of structural models yields biased estimates of, e.g., marginal cost or consumer price elasticity. However, it is also true that specification crucially matters for this type of structural models, as misspecification is likely to affect estimates of agents' preferences or production technology as much as neglecting asymmetric information altogether.

The second approach is more recent in the literature and proposes a way round this problem, in a series of articles based on nonparametric approaches. The move from parametric to nonparametric methods for estimating structural models with asymmetric information was following the development of structural models of auctions, with which they share some common features, as well as the literature on nonparametric identification. (d'Haultfoeuille and Fevrier, 2010 Perrigne and Vuong, 2012).

Structural models of delegated management with asymmetric information share common features with models of auctions or nonlinear pricing. The structural equations are nonlinear in an unobserved heterogeneity component (the private-information parameter) whose distribution is explicitly part of the solution to the underlying economic model. Moreover, endogenous selection of companies may also occur because of individual-rationality or incentive-compatibility constraints, in a way similar to auction participants or consumers in nonlinear pricing problems. For instance, Perrigne and Vuong (2012) propose a nonparametric identification method for models of nonlinear pricing, whose estimating equations closely resemble the ones associated with structural models of management delegation.

We propose in this paper a new methodology for estimating optimal contracts under asymmetric information, which relaxes most assumptions considered in the literature. More precisely, we rely extensively on non parametric techniques to derive the distribution of the private information type and functional estimates of the model. The paper discusses the non parametric identification of the structural model under several assumptions regarding unobserved heterogeneity. The estimation method is applied to the case of delegated management of water utilities in France. The advantage of considering such industry for our empirical ap-

plication lies in the fact that contracts between local communities and private firms for the operation of water utilities in France exist for a long time, under various modes that make the application of the theory of contracts particularly relevant. In the case we consider, the local community manager has incomplete information on the efficiency of the (private) operator of the water utility, before signing the delegation contract. The optimal contract is a second-best solution which depends on marginal social surplus for water, marginal cost of water supply, as well as on the distribution of the private type. A particular aspect of the model is the existence of two sources of unobserved heterogeneity (to the econometrician). Beside the usual private parameter, the econometrician does not observed an heterogeneity term associated with the social surplus and which is community-specific.

The paper is organized as follows. Section 2 presents the French water sector. Section **??** presents a theoretical principal-agent model of contract-based regulation between the local community and the operator of the water utility in a context of information asymmetry. This model is tailored to the specificities of the sector under scrutiny. The system of equations describing the optimal contract are first derived. Those equations summarizes the optimality of behavior of the municipality and the operator. They serve as the basis for the estimation procedure. In Section 4, we discuss the conditions for nonparametric identification of the structural model. Two different cases are considered, depending on assumptions made on unobserved heterogeneity. Section 6 introduces the empirical application to the French water utilities with a particular focus on residential water pricing rules. Some counterfactuals are presented in Section 7. Section 8 discusses alternative modeling of optimal contracting relationships and illustrates some specificities of our approach. Finally, Section 9 concludes.

2   The French Water Sector

2.1   Governance

As in other countries throughout the world, the provision of water services in France is a regulated activity although it has always been fully decentralized at the local level since the 1789 French Revolution. The need to regulation comes from the fact that each stare network is indeed a public monopoly, with high fixed costs and low variable costs, as well as a declining (long-run) average cost curve. In each urban area of significant size equipped with such a water network, a local authority (a single city or a group of cities) has full responsibility to contract with an operator for providing water to the corresponding population.

In terms of size, water utilities represent no less than 1% of French national GDP. Yet,

modes of governance significantly differ across networks. Water utilities in charge of supply, distribution and sewage activities may indeed be public (the so-called *"régies municipales"* or private (*"gestation déléguée"*).[1] Delegation of utility operation to a private company is sometimes viewed as improving efficiency although, empirical studies are not really conclusive as to the relative efficiency of a management mode or the other.[2]

When the service is delegated to the private sector, various kinds of contracts (*"concession"*, *"affermage"*, *"gérance"*..) rule details of this public-private partnership related to water production and distribution, maintenance, and quality supervision.[3] For so-called *régies intéressées*, the operator does not own the network and is paid as a fraction of the benefits of the service that accrues to the municipality. For a contract d' *"affermage"* , the operator is again paid directly by consumers but the cost of maintaining and renewing assets remains borne by the municipality. If the contract is a lease contracts *"concession"*, the operator must cover the costs of building and maintaining the infrastructure which are given back to the municipality at the end of the contract. The operator is directly paid through consumers' bills (the so-called *"redevance"*). By 2007, although private arrangements only concerned 39 % of water services, its share presented 72 % of the overall population. By 2008, 28% of networks were under public management. At the same time, the private sector is highly concentrated with three large companies sharing the market: Veolia Eau France, 39%, Lyonnaise des Eaux, 19 %, Saur, 11% and others companies or joint agreements among big ones amounting to less than 3 %. Veolia is the leading firm in the private sector. The smallest company Saur is more present in rural areas. It is also interesting to notice that public management is more frequent in small municipalities.

Given the high concentration in the sector, practitioners, medias and even politicians have sometimes complained that the market may not be so competitive after all, leaving inexpert municipalities in a weak bargaining position in front of big private players and subject to corruptive behavior. As a result, the *Loi Sapin* was enacted in 1993 to improve transparence and competitive biding while the 1995 *Loi Mazeaud* was designed to improve control on the operator. The competitive bidding generally relies on a two-step procedure. First the public authority chooses an operator through tender, whose criteria need not be publicized. Then a winner is selected after negotiation within those operators having made the best offers.

---

[1]As a matter of comparison, ownership is mostly public in Germany and Italy, while it remains mixed in Spain. In Great Britain, water utilities are always run by private companies under the aegis of an independent national regulator (OFWAT). In the U.S., water utilities operate under rate-of-return regulation. The fact that State commission for public utilities may differ in their mandates gives rise to fairly heterogeneous rates of return.

[2]See Bhattacharyya et al. (1995), Bhattacharyya et al. (1994) and Estache and Rossi (1999) among others.

[3]In France, the public authority itself is in charge of verifying the quality of the service.

2.2 Pricing

Importantly, heterogeneity is a key aspect of the sector both on the demand and the cost sides. Networks are disconnected from one municipality to the other and vary significantly in terms of length, age, and thus leakages. For instance, although leakages amount only to 3% in Paris, they may reach up to 40% in some rural areas to average overall 21,9%. More maintenance helps reducing leakages and save water resource. Even though existing infrastructures are often old, the replacement rate for years 2006-2008 remained rather small with only 0,6% of the networks being renewed over that period.[4]

Given the specificities of each network, the operator managing the service may learn over time the state of the network. Private information over the cost of providing the service is thus pervasive. It introduces a fundamental asymmetry between municipalities and operators. This asymmetry is also partly due to the difficulties in assessing how costs, including labor costs, overhead costs, and maintenance investments, are allocated between water supply, distribution or sewage treatment, especially when more than one of these three operations are shared by the same operator.[5]

Such heterogeneity explains also significant disparities in consumption prices, with an average bill of 183 euros per inhabitant by 2008 which slightly increases over time as a mean to finance an improving quality or new investments.

As far as pricing is concerned, water is billed with two-part tariffs; an usual feature of pricing in network industries.[6] The fixed-fee helps to cover fixed cost while the variable part that depends on consumption aims at paying for variable costs. Fixed-fees significantly vary across networks.

3   Theory

Our theoretical model of the contractual relationship between municipalities (sometimes referred to as "principals" in the sequel) and service providers (the "agent') fits the actual contractual practices reviewed in Section 2. In particular, informational asymmetries, heterogeneity and the form of pricing are key ingredients of any formal description of the sector.

---

[4]See Commissariat Général au Developpement Durable (2010).

[5]For a discussion of cost in the French stare sector, see Garcia and Thomas (2001).

[6]Hall (2000).

3.1 Preliminaries

We first set up the stage before entering into more details into the characterization of an optimal contract in the environment under scrutiny.

Demand side. We consider a population of heterogeneous municipalities which differ in terms of the consumers surplus that prevails locally. More precisely, we assume that there exists a shift parameter ε, common knowledge for contracting parties (although not observed by the econometrician), such that the surplus in a municipality characterized by ε writes as $S(q, \varepsilon)$, where $S(q, \varepsilon)$ is increasing and concave in the consumption q and increasing in ε. This parameter ε allows to take into account heterogeneity on the demand side.

In a given municipality, aggregate demand for water at price p is then denoted as $D(p, \varepsilon) = (S'_q)^{-1}(p, \varepsilon)$ with $D'_p(p, \varepsilon) = \frac{1}{S''_{qq}(D(p,\varepsilon),\varepsilon)} < 0$) for any realization of ε. Of course, the following identity holds

$$p = S'_q(D(p, \varepsilon), \varepsilon).$$

Supply side. The cost function of the service operator is parameterized as $\theta C_0(q)$, where q is the amount produced and and θ is an efficiency parameter that enters multiplicatively. We assume that the function C_0 is strictly increasing and convex. Observe that this cost function satisfies the usual Spence-Mirrlees assumption; an operator with a more efficient technology (θ lower) also produces at a lower marginal cost. As usual in the screening literature, this assumption ensures that different operators can be sorted according to their marginal cost of producing the service and choose accordingly to produce under different contractual terms.

Information. In this paper, we are instead interested in the case where the cost parameter θ is the firm's private information. This parameter is distributed according to a common knowledge atomless distribution F, with a positive density function f on a bounded support $\Theta = [\underline{\theta}, \bar{\theta}]$. In accordance with the screening literature,[7] we impose the familiar *monotone hazard rate property*[8] that ensures fully separation allocations at an optimal contract.

Assumption 1 (MHR).
$$\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) \geq 0, \quad \forall \theta \in \Theta.$$

---

[7]Guesnerie and Laffont (1984) and Laffont and Martimort (2002, Chapter 3).
[8]See Bagnoli and Bergstrom (2005).

Contracts. The contract between the municipality and the firm stipulates not only a price p per unit of water produced but also an upfront subsidy A under the form of (per capita) subscription fees paid by consumers to access the service. That subsidy distribute the overall surplus between the consumers and the operator. Implicit in this specification of the contract is the idea that controlling the unit price amounts to controlling demand and thus the production that meets this demand.[9]

Following the incentive regulation literature, we will envision the result of contractual negotiations between the firm and the municipality as the choice of an item by the privately informed party within a menu of options. Two equivalent approaches might be used to model this choice. The first one relies on the so-called *Revelation Principle*[10] which states that there is no loss of generality in looking for contracts which are direct and truthful mechanisms of the kind $\{A(\hat{\theta}, \varepsilon), p(\hat{\theta}, \varepsilon)\}_{\hat{\theta} \in \Theta}$. Note that we index the contract by the demand shock ε which is commonly known by contracting parties. With such direct communication, the operator picks a subsidy/unit price price according to his efficiency parameter by communicating information on its cost parameter. The mechanism is incentive compatible when each operator ends up preferring the option targeted to his own type.

An alternative approach based on the so-called *Taxation Principle*[11] gives up the abstract direct communication process underlying the Revelation Principle and focuses instead on the true economic choice made by the privately informed party. Facing a nonlinear scheme $A(p, \varepsilon)$[12] linking the value of the subscription fee to the actual per unit price chosen by the firm, the firm chooses optimally at which price it stands to produce.[13] Again, this choice can be viewed as a metaphor for more complex negotiation procedures where firms and public authorities negotiate over both the fixed fee and the unit price charged to consumers. We will favor this second approach since it relates to actual observables available to us (price, quantity, fees) which will be used in our econometric analysis.

Objective functions of the contracting parties. With our previous notations at hands, the

---

[9]See Baron (1989a) for a general formulation of regulatory mechanisms relying on such approach. It is only incidentally different from the standard approach that focuses on the direct control of quantities that is developed in Baron and Myerson (1982) and Laffont and Tirole (1993).

[10]See Myerson (1982) and Laffont and Martimort (2002) for a textbook approach.

[11]See for instance Rochet (1985).

[12]Without fear of confusion and for the sake of simplifying presentation, we slightly abuse notations here by denoting similarly the fixed fee viewed as a function of the agent's type in a direct mechanism and viewed as a function of the price in an indirect scheme. Notice also that the nonlinear scheme $A(\cdot, \varepsilon)$ is again indexed on the commonly observable variable ε that characterizes the relationship under scrutiny.

[13]Of course, any incentive compatible direct mechanism $\{A(\hat{\theta}, \varepsilon), p(\hat{\theta}, \varepsilon)\}_{\hat{\theta} \in \Theta}$ can be transformed into a nonlinear scheme by setting $A(p, \varepsilon) = A(\hat{\theta}, \varepsilon)$ if $p = p(\hat{\theta}, \varepsilon)$ and $A(p, \varepsilon) = -\infty$ otherwise.

8

expression of the firm's profit becomes:

$$\mathcal{U}(\theta, \varepsilon, p, A) = A + pD(p, \varepsilon) - \theta C_0(D(p, \varepsilon)).$$

Following Baron and Myerson (1982), a municipality maximizes a welfare function which includes not only consumers' net surplus from consuming the service but also the firm's profit weighted by some parameter $\gamma \in [0, 1[$. This parameter can be viewed as an index of the firm's bargaining power at the the of tenders or during contract negotiations. It can also be inherited from how local political forces interact as argued in Baron (1989).[14] The corresponding welfare function $W(\cdot)$ can thus be written as:

$$\mathcal{W}(\theta, \varepsilon, p, A) = S(D(p, \varepsilon), \varepsilon) - A - pD(p, \varepsilon) + \gamma \mathcal{U}(\theta, \varepsilon, p, A).$$

Using the expression of the fixed-fee as a function of the firm's profit, the latter definition becomes:

$$\mathcal{W}(\theta, \varepsilon, p, A) = S(D(p, \varepsilon), \varepsilon) - \theta C_0(D(p, \varepsilon)) - (1 - \gamma) \mathcal{U}(\theta, \varepsilon, p, A).$$

This latter expression stresses the rent/efficiency trade-off faced by the local government in designing the regulatory contract. On the one hand, the principal would like to charge a price $p(\theta, \varepsilon)$ close to $p^*(\theta, \varepsilon)$ as defined in (1) so that the overall surplus is maximized. On the other hand, the principal would also like to reduce the firm's informational rent which is viewed as socially costly. Under asymmetric information, rents and outputs are linked altogether through incentive compatibility conditions and this leads to an important trade-off between the conflicting objectives of promoting efficiency and extracting rents.

Remark 1. *The expressions of the objective functions above can easily be extended to accounts for some fixed-cost F in the operatopr's cost function. Suppose indeed that the operator's profit can be written as $\mathcal{U}(\theta, \varepsilon, p, A) = A' + pD(p, \varepsilon) - \theta C_0(D(p, \varepsilon)) - F$. for some fixed-fee A'. Setting $A \equiv A' - F$ then amounts to having the principal pays for the fixed-cost in the first place, which is basically an accounting convention.*

Benchmark. Had θ been common knowledge, efficiency would require to produce a quantity $q^*(\theta, \varepsilon) = D(p^*(\theta, \varepsilon), \varepsilon)$ such that the marginal social value of production is equal to marginal cost:

$$S_q'(q^*(\theta, \varepsilon), \varepsilon) = p^*(\theta, \varepsilon) = \theta C_{0q}'(q^*(\theta, \varepsilon)). \tag{1}$$

---

[14]See also Gagnepain, Ivaldi and Martimort (2013) for a model of the French transportation sector that relies on a similar specification of the preferences of local authorities.

Then, the firm's operates if $A^*(\theta, \varepsilon)$ extracts revenues from the service:

$$A^*(\theta, \varepsilon) = -p^*(\theta, \varepsilon)D(p^*(\theta, \varepsilon), \varepsilon) + \theta C_0(D(p^*(\theta, \varepsilon), \varepsilon)). \tag{2}$$

Remark 2. *In the empirical part of our analysis, the two functions $S(\cdot, \varepsilon)$ and $C_0(\cdot)$ will depend on a set of explanatory variables. For example, the treatment made for making the water drinkable has an impact on the unit cost. This treatment is however observed in the data. For the exposition, we omit the dependance in the explanatory variables without loss of generality for the results derived in the analysis of the theoretical model.*

3.2 Optimal Contract

Incentive compatibility constraints. Let define the firm's information rent $U(\theta, \varepsilon)$ and an optimal price[15] respectively as:

$$U(\theta, \varepsilon) = \max_p A(p, \varepsilon) + pD(p, \varepsilon) - \theta C_0(D(p, \varepsilon)) \tag{3}$$

and

$$p(\theta, \varepsilon) = \arg\max_p A(p, \varepsilon) + pD(p, \varepsilon) - \theta C_0(D(p, \varepsilon)). \tag{4}$$

From (3), $U(\theta, \varepsilon)$ is the maximum of a family of decreasing linear functions in θ. As such it is decreasing, convex in θ and absolutely continuous so that one can write:

$$U(\theta, \varepsilon) = U(\bar{\theta}, \varepsilon) + \int_\theta^{\bar{\theta}} C_0\left(D(p(x, \varepsilon), \varepsilon)\right) dx. \tag{5}$$

At any point of differentiability in θ (i.e., almost everywhere), we get:

$$U'_\theta(\theta, \varepsilon) = -C_0\left(D(p(\theta, \varepsilon), \varepsilon)\right). \tag{6}$$

Because $U(\theta, \varepsilon)$ is convex in θ, $C_0\left(D(p(\theta, \varepsilon), \varepsilon)\right)$ is non-decreasing in θ, which in turn implies:

$$p(\theta, \varepsilon) \text{ is non-increasing in } \theta. \tag{7}$$

This condition expresses the fact that firms endowed with less efficient technologies produce lower volumes at higher prices . From this monotonicity, it also follows that $p(\theta, \varepsilon)$ is almost everywhere differentiable in θ.

---

[15]Or at least, a selection within the best-response correspondence.

Let us turn now to the expression of the nonlinear schedule $A(p, \varepsilon)$ that plays an important role in our empirical analysis. By a standard duality argument of (*generalized*) convex analysis, we may first rewrite:[16]

$$A(p, \varepsilon) + pD(p, \varepsilon) = \min_{\theta} U(\theta, \varepsilon) + \theta C_0(D(p, \varepsilon)).$$

From which, it also follows that $A(p, \varepsilon)$ is absolutely continuous in p and thus such that

$$
\begin{aligned}
A(p, \varepsilon) + pD(p, \varepsilon) = & A(p(\bar{\theta}, \varepsilon), \varepsilon) + p(\bar{\theta}, \varepsilon)D(p(\bar{\theta}, \varepsilon), \varepsilon) \\
& + \int_{p(\bar{\theta}, \varepsilon)}^{p} \vartheta(p, \varepsilon)C_{0q}' (D(p, \varepsilon)) D_p'(p, \varepsilon)dp
\end{aligned}
\tag{8}
$$

where $\vartheta(p, \varepsilon) = \min_{\theta} U(\theta, \varepsilon) + \theta C_0(D(p, \varepsilon))$ is an assignment function[17] (a selection within the best-response monotonically increasing p (thus almost everywhere differentiable) and such that $\vartheta(p(\theta, \varepsilon), \varepsilon) \equiv \theta$. At any point of differentiability in $p = p(\theta, \varepsilon)$, we thus have also:

$$A_p'(p(\theta, \varepsilon), \varepsilon) = -\left(p(\theta, \varepsilon) - \theta C_{0q}' (D(p(\theta, \varepsilon), \varepsilon)) \right) D_p'(p(\theta, \varepsilon), \varepsilon) - D(p(\theta, \varepsilon), \varepsilon). \tag{9}$$

Participation constraints. The firm chooses to always operate the service irrespectively of its costs when it at least breaks even. For a fixed ε, this participation constraint can be written as:

$$U(\theta, \varepsilon) \geq 0 \qquad \forall \theta.$$

As usual in the incentive regulation literature,[18] this constraint is binding for the worst type $\bar{\theta}$ at the optimal contract. Otherwise reducing uniformly the fixed fee by some small amount would improve the principal's expected payoff while maintaining incentive compatibility. From this observation, an immediate manipulation of (5) yields the following expression of the firm's information rent as:

$$U(\theta, \varepsilon) = \int_{\theta}^{\bar{\theta}} C_0 (D(p(x, \varepsilon), \varepsilon) dx. \tag{10}$$

Observe that the rent left to the operator is greater as prices are lower. The intuition is a standard one. By pretending being slightly less efficient, an operator with efficiency parameter θ can produce the same quantity than this slightly less efficient type $\theta + d\theta$ but

---

[16]See for instance Basov (2005, Chapter 7).

[17]See Noldecke and Samuelson (2007).

[18]See Armstrong and Sappington (2007), Baron and Myerson (1982) and Laffont and Tirole (1993) among others.

at a lower marginal cost. To induce this operator to report truthfully his type he must be given an extra fee that equals the corresponding cost saving $d\theta C_0 \left(D(p(\theta + d\theta, \varepsilon), \varepsilon) \right) \approx d\theta C_0 \left(D(p(\theta + d\theta, \varepsilon), \varepsilon) \right)$. The right-hand side of (10) expresses how those marginal information rents just pill up over all supra-marginal types.

Optimal contracts. Under asymmetric information, an optimal contract maximizes the expected welfare of the municipality subject to incentive and participation constraints. From our observations above, that incentive feasible set can be summarized by constraints (7) and (10). As usual, the monotonicity condition (7) will be omitted in a first step and checked ex post on the solution to the so relaxed problem. Formally, this relaxed problem can be written as:

$$\max_{\{p(\cdot,\varepsilon),U(\cdot,\varepsilon)\}} \int_{\underline{\theta}}^{\overline{\theta}} \left[S(D(p(\theta, \varepsilon), \varepsilon)) - \theta C_0 \left(D(p(\theta, \varepsilon), \varepsilon) \right) - (1-\gamma)U(\theta, \varepsilon) \right] dF(\theta) \text{ subject to (10).}$$

Using (10) and integrating by parts yields the following expression of the expected rent left to the operator:

$$\int_{\underline{\theta}}^{\overline{\theta}} U(\theta, \varepsilon) dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} \frac{F(\theta)}{f(\theta)} C_0 \left(D(p(\theta, \varepsilon), \varepsilon) \right) dF(\theta).$$

This expression can be incorporated into the maximand above before proceeding to point-wise optimization. This last step leads to the following expression of the price per-unit of consumption $p(\theta, \varepsilon)$.

$$p(\theta, \varepsilon) = \left(\theta + (1-\gamma)\frac{F(\theta)}{f(\theta)} \right) C'_{0q} \left(D(p(\theta, \varepsilon), \varepsilon) \right). \tag{11}$$

The corresponding volume that is supplied $q(\theta, \varepsilon)$ is then defined as:

$$S'_q(q(\theta, \varepsilon), \varepsilon) = p(\theta, \varepsilon). \tag{12}$$

Equation (11) indicates that the price is now above marginal costs and, as a result of (12). Equilibrium quantities are also lower than at the first best. Increasing the unit price above marginal cost reduces the demand addressed to the operator. It thus reduces the latter's information rent. Formally, everything happens as if the cost parameter was now replaced by a *virtual cost parameter* $H(\theta, \gamma)$ which is greater:

$$H(\theta, \gamma) = \theta + (1-\gamma)\frac{F(\theta)}{f(\theta)}.$$

This expression first illustrates the rent/efficiency trade-off that arises under asymmetric information and, second, how this trade-off is modified as parameters of the model change. Indeed, the virtual cost parameter is greater as the public authority is more concerned by rent extraction (i.e., γ lower) and as the types distribution is more front-loaded, in the sense of having a greater hazard rate $F(\theta)/f(\theta)$.

Turning now to sufficient conditions for optimality, observe that Assumption 1 ensures that $p(\theta, \varepsilon)$ so defined by (11) is non-decreasing in θ as requested by condition (7). Hence, the solution of the relaxed problem really characterizes the optimal contract. When we get to our empirical analysis, we will actually check on our estimated distribution that it indeed satisfies Assumption 1.

4   Nonparametric Identification

In this section, we study the nonparametric identification of our model. It is indeed important to figure what kind of structural functions which can be fully determined by the available data. Identification is related to the variables observed by the econometrician which in turn depends on the set of variables used for contracting purposes. In our dataset, we observe for each local community, the unit price p, the quantity consumed q and the fixed fee A. We also observe some explanatory variables W and Z which are related to respectively the cost function $C_0(q, W)$ and the surplus function $S(q, \varepsilon, Z)$. The vectors W and Z could have variables in common but we assume that we have exclusion restrictions, i.e., there exists an explanatory variable in Z which is not part of W and an explanatory variable in W which is not part of Z. Our purpose is to identify the production technology $C_0(q, W)$, the distribution $F(\theta)$ of the private types θ, the consumers surplus function $S_0(q, \varepsilon, Z)$, and the weight γ of the firm's profit in the principal's objective.

As a matter of fact, it is straightforward to notice than multiplying θ by a positive scalar β^0 and dividing the cost function by the same value would give the same equilibrium outcome. This observation due to Perrigne and Vuong (2012) is by now pretty standard in this literature. It follows that we shall normalize for the distribution of θ. by assuming that the lowest value of θ, $\underline{\theta}$, is equal to 1. The function $C'_{0q}(q, W)$ is therefore interpreted as the marginal cost function for the most efficient type, given the observed characteristics W.

Assumption 2 (Normalization)**.**

$$\underline{\theta} = 1.$$

4.1 The Simple Case Without Heterogeneity

To start with, we do not consider neither the presence of explanatory variables (no W and no Z) nor the presence of heterogeneity, *i.e.* $\varepsilon \equiv 0$, simplifying notations accordingly. For exposition purpose, it is important to know what are the identifying power of the three equilibrium equations (9), (11) and (12) and what do the additional assumptions help to identify. The system of first-order conditions (11)-(12) and (9) reduces in this case to[19]

$$p = H(\theta, \gamma) C'_{0q}(q), \tag{13}$$

$$S'(q) = p, \tag{14}$$

$$A'(p) = -q - \left(p - \theta C'_{0q}(q)\right) D'_p(p). \tag{15}$$

Observe that equation (14) directly identifies $S'(q)$ on the support of the equilibrium quantities and hence provides the expression of $D'_p(p) = \frac{\partial}{\partial p}\left(S'^{-1}\right)(p)$ that we insert into (15). Finally, (15) provides information about the operator's price-cost margin

$$\frac{p - \theta C'_{0q}(q)}{p} = \frac{A'(p) + q}{p D'_p(p)}.$$

This is so because the terms on the right-hand side are either observed ($q = D(p)$ and p) or derived directly from the observations ($A'(p)$ and $D'(p)$).

Making the dependence of the price-cost margin on θ explicit, we define a price-cost margin $r(\theta, \gamma)$ as:

$$r(\theta, \gamma) = \frac{p(\theta) - \theta C'_{0q}(q(\theta))}{p(\theta)} = \frac{1}{1 + \frac{\theta f(\theta)}{(1-\gamma)F(\theta)}} \tag{16}$$

where the last equality immediately follows from (14). Thus, the price-cost margin only depends only on the efficiency parameter θ and the bargaining power γ. In the sequel, we shall assume that different operators can be perfectly sorted according to that price-cost margin. Formally, we require that there is a one-to-one mapping between price-cost margins and efficiency parameters, i.e., $r(\theta, \gamma)$ is a monotonically decreasing transformation of θ which always lies between 0 and 1 and is worth 0 at $\underline{\theta}$, i.e., for the most efficient type who produces efficiently:

Assumption 3 (MPCM)**.**

$$\frac{d}{d\theta}\left(\frac{F(\theta)}{\theta f(\theta)}\right) \geq 0, \quad \forall \theta \in \Theta.$$

---

[19]We omit, in the notations, the dependance of p and q in θ for the clarity of exposition.

This assumption is actually weaker than Assumption 1 and is satisfied for all standard parametric distributions.

To give a bit more intuition about the role played by Assumption 3, let us come back to (15) which can be rewritten as:

$$A'(p(\theta)) = -D(p(\theta))(1 + r(\theta)\varepsilon_D(p(\theta))) \qquad (17)$$

where we denote the demand elasticity by $\varepsilon_D(p) = -\frac{pD'(p)}{D(p)}$. Assuming also that demand is more elastic at greater price, i.e., $\varepsilon_D(p)$ is increasing with p, it is straightforward to check by differentiating (17) that Assumption 3 ensures that $-A'(p(\theta))/D(p(\theta))$ is increasing in θ, or alternatively that $A(p)$ is quasi-concave in p.

Equipped with this one-to-one relationship between cost-price margins and efficiency parameters, we now let $G(\cdot)$ (resp. $g(\cdot)$) be the cumulative distribution function (resp. probability density function) of that margin. Of course, we have $G(r(\theta, \gamma)) = F(\theta)$ and by differencing the last equality with respect to θ, we also get $g(r) = f(\theta)\theta'_r(r, \gamma)$ where $\theta(r, \gamma)$ denotes the inverse function of $r(\theta, \gamma)$. Reintroducing the last two equalities into (16), we obtain after some manipulations:

$$(1 - \gamma)\frac{\theta_r(r, \gamma)}{\theta(r, \gamma)} = \frac{r}{1 - r}\frac{g(r)}{G(r)}.$$

As, $r = 0$ when $\theta = \underline{\theta} = 1$, we can solve the latter differential equation and get:

$$\theta(r, \gamma)^{1-\gamma} = \exp\left[\int_0^r \frac{s}{1 - s}\frac{g(s)}{G(s)}ds\right]. \qquad (18)$$

The density function for θ can then be easily derived from $f(\theta) = \frac{g(r)}{\theta(r)}$:

$$f(\theta) = (1 - \gamma)\frac{1 - r}{r\theta}G(r). \qquad (19)$$

Observe that the last expression derives the density of the types as a function of the c.d.f. of the price cost margin. Reintroducing (18) into (13) leads to

$$c'_{0q}(q) = \frac{p}{H(\theta, \gamma)} = \frac{p(1 - r)}{\exp\left[\frac{1}{1-\gamma}\int_0^r \frac{s}{1-s}\frac{g(s)}{G(s)}ds\right]}.$$

Therefore, when γ is known or predetermined from some additional source of information, the model is thus identified since the marginal cost and the distribution of the firms' types are identified up to the standard normalization.

15

Instead, when γ is not known, only $\theta^{1-\gamma}$ is identified. Consequently, for any β in $[0,1[$, the two sets $(\theta, c'_{0q}(q), \gamma)$ and $(\theta^{\frac{1-\gamma}{1-\beta}}, c'_{0q}(q)\theta^{\frac{\gamma-\beta}{1-\beta}}, \beta)$ are observationally equivalent. We thus need additional assumptions or information to identify the model. It is worth noting that the conclusions of this section does not depend on the existence of observed explanatory variables. Adding an explanatory variable in either the cost function or the surplus one does not change at all the previous results. Observe also that we can exploit the bounds on β to bound the distributions of interest. In particular

$$c'_{0q}(q) \leq \frac{p}{H(\theta, \gamma)} = \frac{p(1-r)}{\exp\left[\int_0^r \frac{s}{1-s}\frac{g(s)}{G(s)}ds\right]},$$

and

$$f(\theta) \leq \frac{1-r}{r\theta}G(r), \quad \text{on } [1; +\infty[.$$

4.2 Full Identification of the Model with Explanatory Variables and Heterogeneity

We now consider the full model that is estimated in Section **??**.

4.2.1 Identification of the Marginal Surplus Function under Completeness Assumption

• In a first step, we assume that the marginal cost function is separable and additive in both the explanatory variables, Z, and the unobserved heterogeneity term, ε, i.e. $S'_q(q, \varepsilon, Z) = S'_0(q) + \beta Z + \varepsilon$, Equation (12) can be written

$$p = S'_0(q) + \beta Z + \varepsilon, \tag{20}$$

with $[\varepsilon|Z, W] = 0$.

The equilibrium quantity q being endogenous, equation (20) is a typical non-parametric instrumental regression like the one studied in Florens et al. (2005).[20] Following Therorem 3 in Florens, Johannes and van Bellegem (2005), we recover the identification of the primitives of the model under a condition of completeness of $\tilde{W}$, the variables in W which are not in Z with respect to q. Observe that here, in the partial linear regression like the one in (20), we need additional regularity conditions than the completeness assumption due to the linear

---

[20]See also Newey and Powell (2003), Blundell et al. (2007), Darolles et al. (2011) and Ai and Chen (2003) for as additional references.

16

part βZ (see Florens et al., 2005)[21]. The completeness assumption is nevertheless the most restrictive assumption though quite standard in this literature. The idea of completeness is that the instruments $\tilde{W}$ are sufficiently rich to identify any function of q:

The set of instruments $\tilde{W}$ is complete for q if, for any measurable function ΔS in L^1,

$$E[\Delta S(q)|\tilde{W}] = 0 \; a.s. \; \Rightarrow \Delta S(q) = 0 \; a.s.$$

Sufficient conditions for completeness can be found in Newey and Powell (2003), Chernozhukov, Imbens, and Newey (2007), and Andrews (2011). It can be replaced by weaker concepts like the bounded completeness in Chernozhukov and Hansen (2005), Blundell, Chen and Kristensen (2007), D'Haultfoeuille (2010), and Andrews (2011).[22]

- We now consider the non-separable case,

$$p = S'_q(q, \varepsilon, Z),$$

with $[\varepsilon|Z, W] = 0$. Let V be indeed a given c.d.f., *i.e* an injective function from $\mathbb{R}$ to $[0, 1]$. Let $\tilde{\varepsilon} = V^{-1} \circ \Phi(\varepsilon)$, where $\Phi(\dot{)}$ denotes the c.d.f. of ε, and $\tilde{S}(q, \tilde{\varepsilon}, Z) = S(q, \Phi^{-1} \circ V(\tilde{\varepsilon}), Z)$. The equilibrium quantities A, p and q are invariant to this transformation.

The distribution of the parameter ε is therefore non identifiable as such and we need a normalization, like for the surplus function. We therefore need an extra normalization assumption which is standard in this literature: $\varepsilon \sim U[0; 1]$. It is worth noting, that this normalization does not restrict the economic interpretation of our model. The value of ε does not have an interpretation in its own and is only an index of the position of the firm in a demand ranking. We therefore identify the marginal surplus (given Z) of any quantile of the distribution of ε, like the marginal surplus of the median local community, given Z.

The identification arises from the monotonicity constraints implied by the economic model (see, for example Chesher, 2007, or Chernozhukov and Hansen, 2005)For a given rank $\alpha \in [0, 1]$:

$$\mathbb{P}\left(p \leq \left(S'_q(q, \alpha, Z)\right) | Z, W\right) = \mathbb{E}_{q|Z,W} \mathbb{P}\left(S'_q(q, \varepsilon, Z) \leq \left(S'_q(q, \alpha, Z)\right) | q, Z, W\right)$$
$$= \mathbb{E}_{q|Z,W} \mathbb{P}(\varepsilon \leq \alpha | q, Z, W) \text{ due to the monotonicity in } \varepsilon \quad (21)$$
$$= \mathbb{P}(\varepsilon \leq \alpha | Z, W) = \alpha$$

---

[21]It is required that the conditional expectations of Z, q and any L^2 function of q given $\tilde{W}$ are in $L^2(\mathbb{R}^{dim\tilde{W}})$ and that the matrix $E\left[E(Z|\tilde{W})E(Z|\tilde{W})^\top\right]$ is full rank.

[22]See also Chen et al. for a general discussion. This assumption is however not testable as recently shown by Canay et al. (2012) .

Equation (21) derived above can be reexpressed as a conditional moment restriction:

$$\mathbb{E}\left(\mathbf{1}\{p \le S'_q(q, \alpha, Z)\} - \alpha | Z, W\right) = 0. \tag{22}$$

This is a standard quantile IV equation and we need essentially an assumption of completeness of $\tilde{W}$, the variables in W which are not part of Z, with respect to q to identify any quantile of the marginal surplus function given Z.The following result summarizes the main conclusion if this section.

Proposition 1. *If $\tilde{W}$ is complete with respect to q the marginal surplus function given Z, $S'_0(q, Z)$, is identified.*

4.3   Identification of the Marginal Cost Function

We now prove that the marginal cost function is identified. Assume initially that we do not have explanatory variables for the marginal cost. For the most efficient firms, $\theta = \underline{\theta}$, and the marginal price equation (11) shows that the marginal cost function $C'_{0q}(q)$ can be estimated using the lower envelope of the points in the space (q, p) between the quantities q^l_0 and q^u_0. q^l_0 is the quantity which is related to the points (q, p) with the lowest price, i.e. the contract for the most efficient firm contracting with the local community with the lowest marginal surplus. q^u_0 is the maximum quantity in the set of contracts, i.e. the quantity in the contract of the most efficient firm contracting with the local municipality with the highest marginal surplus function.

Insert Figure here

Figure 1: The set of contracts in the $(q \times p)$ space.

For the less efficient firms, $\theta = \overline{\theta}$ and the marginal prices are equal to $kC'_{0q}(q)$ where k is equal to the limit of $\theta + (1 - \gamma)/f(\theta)$ on the upper bound, if it exists. This function is

18

therefore the upper envelope between q_1^l (the minimum quantity among all contracts i.e. the quantity of the contract of the less efficient firm contracting with the local community with the lowest surplus) and q_1^u (the quantity consumed in the local community with the highest surplus which contracts with the less efficient firm).

If $q_1^u \geq q_0^l$ and if k is finite, we can estimate k by the ratio of two extreme costs with the same quantity q and therefore estimate $C_{0q}'(q)$ between q_0^l and q_1^u. We need to have sufficient heterogeneity in ε. If there are explanatory variables W, we can use the same argument conditionally on W.

Alternatively, if k is infinite, or if $q_1^u < q_0^l$, we can use a moment equation like the one used for estimating the surplus function in (21). Using the monotone property of $H(\theta) = \theta + (1 - \gamma)\frac{F(\theta)}{f(\theta)}$ in θ, we can indeed, write:

$$\mathbb{P}\left(p \leq H(\alpha)C_{0q}'(q, W)|Z, W\right) = \alpha. \tag{23}$$

For any α, we can identify both $H(\alpha)$ and $C_{0q}'(q)$ under an assumption of completeness of $\check{Z}$, the explanatory variables which are part of Z but not part of W with respect to q. The result is summarized in the next proposition.

Proposition 2. *If $\check{Z}$ is complete with respect to q the marginal cost function given W, $C_{0q}'(q, W)$ and $H(\theta) = \theta + (1 - \gamma)\frac{F(\theta)}{f(\theta)}$ are identified.*

4.4   Identification of θ and γ

We now end the identification process by focusing on γ and θ.

Proposition 3. *If $\check{Z}$ and $\tilde{W}$ are complete with respect to q, the bargaining power of the firm γ and the inefficiency types θ are identified.*

Once we know, $C_{0q}'(q, W)$, we can recover $H(\theta)$ for any quantile of the distribution of θ from (11). The cost parameter θ can therefore be derived from (9), as all remaining quantities are known or derived from the previous steps. The knowledge of θ implies the knowledge of both its c.d.f. and p.d.f. Consequently, the bargaining weight γ is identified.

4.5   Comments about the Role of Heterogeneity

To be done

19

5 Simulations and Estimation Procedure

In this section, we present the different steps of our estimation strategy. Some Monte Carlo simulations are also displayed to assess the sample properties of the proposed estimator.

5.1 Estimation of the marginal surplus function

We assume that the marginal surplus is separable and additive in both the explanatory variables, Z and the unobserved heterogeneity ε like in Equation (20):

$$p = S'_q(q) + \beta_Z Z + \varepsilon. \tag{24}$$

We apply Ai and Chen (2003) to estimate the marginal surplus using cubic spline approximation. This is relatively easy to implement and the decreasing property of $S'_q(q)$ can be forced by imposing that the sequence of coefficients related to the spline approximation is decreasing:

1. We choose the cubic splines $B_4(t)$ as basis for our sieve estimation and renormalize q on $[0, 1]$ for the choice of the knots (G denotes the c.d.f. of q):

$$S'_q \circ G^{-1}(q^*) = \phi(q^*) = \sum_{l=-3}^{2^k-1} B_4(2^k q^* - l)$$
$$= P_{q^*} \beta^{K_n},$$

   where $K_n = 2_n^d + 4$. d_n is a tuning parameter which drives the approximation of the true (normalized) marginal surplus function.

2. We then project the moment equation

$$m(Z, W, \beta^{K_n}, \beta_Z) = \mathbb{E}\left[p - P_q \beta^{K_n} - \beta_Z Z | Z, W\right]$$

   on a basis of the square-integrable function of Z, W, $p_j(Z, W)$, $j = 1...J_n$, where J_n tends slowly to infinity as $n \to \infty$. We denote by $P^{J_n}(Z, W) = (p_1(Z, W), ..., p_{J_n}(Z, W))$ and $P = (P^{J_n}(Z_1, W_1), \dots, P^{J_n}(Z_n, W_n))$. An empirical estimator of $m(Z, W, \beta^{K_n}, \beta_Z)$ is therefore:

$$\hat{m}(Z, W, \beta^{K_n}, \beta_Z) = \sum_{j=1}^n \left(p_j - P_{q^*_j} \beta^{K_n} - \beta_Z Z_j\right) P^{J_n}(Z_j, W_j) \left(P'P\right)^- P^{J_n}(Z, W).$$

We can now compute the (empirical) distance function $\hat{Q}(\beta_Z, \beta^{K_n})$ as:

$$\hat{Q}(\beta_Z, \beta^{K_n}) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(Z_i, W_i, \beta^{K_n}, \beta_Z) \hat{m}(Z_i, W_i, \beta^{K_n}, \beta_Z).$$

We could have used another metric but it appears that the result are not so sensitive to the variance matrix that could have been used to estimate the distance function

3. Finally, we estimate the parameters β_Z and β_{K_n} by minimizing the penalized criterion:

$$\hat{Q}(\beta_Z, \beta^{K_n}) + \kappa_n(d_0 + d_2).$$

The second term in the expression above is a penalization term added to control for the ill-posedness where $d_0 = \int_0^1 \phi(q^*)^2 dq^*$ and $d_2 = \int_0^1 \phi''(q^*)^2 dq^*$. d_2 is controlling for the oscillations of the estimated marginal surplus. κ_n is a tuning parameter which controls for the strength of the penalisation term.

5.2 Estimation of the marginal cost function

We estimate the marginal cost function using the SMD procedure of Chen and Pouzo (2011) which is, among additional contributions, a generalization of Ai and Chen (2003) to the case of nonsmooth moments. It is therefore particularly attractive here as we estimate the marginal cost function by quantile IV methods. The marginal cost function $C'_{0q}(q)$ is estimated by another spline approximation where its monotonicity is imposed in the etimation procedure. This is otherwise the same steps than the ones related to the estimation of the marginal surplus function where the moment condition $\mathbb{E}\left[p - P_q \beta^{K_n} - \beta_Z Z | Z, W\right] = 0$ is replaced by

$$\mathbb{E}\left[\mathbf{1}\{\left(p \leq (\lambda_\alpha + \beta_W W) P_{q^*} \delta^{J_n}\right)\} - \alpha | Z, W\right] = 0 \text{ ,for any } \alpha \in]0, 1[. \quad \lambda_\alpha \text{ is an increasing}$$
sequence of parameter which corresponds to the value $H(\theta_\alpha, \gamma)$ for the α-quantile of the firm type, θ_α.

In practice, we will select a few quantiles to estimate $C'_{0q}(q)$ as we can directly recover $H(\theta, \gamma)$ by computing the ratio of the prices to the marginal costs. Finally, Chen and Pouzo proposes a relatively simple wild bootstrap technique that is valid in the context of quantile IV which validates our empirical strategy for estimating confidence intervals.

21

5.3 Estimation of the types and of the weight parameter γ

From the estimation of the marginal cost, we can now estimate $H(\theta, \gamma)$ in (11) by the ratio of the price to the marginal cost. Conditional on γ,

$$H(\theta, \gamma) = \theta + (1 - \gamma)\frac{F(\theta)}{f(\theta)}.$$

As $H(\theta)$ is a monotonic transformation of θ, we can recover θ by inverting the previous relationship using the boundary condition $H(\underline{\theta}) = H(1) = 1$:

$$\theta(\gamma) = H(\theta, \gamma) - \frac{1}{K(H(\theta, \gamma))} \int_1^{H(\theta, \gamma)} K(u)du,$$

where $K(\cdot)$ is the c.d.f. of $H(\theta, \gamma)$. This equality is common on the auction literature (see Guerre et al. (2002) among others).

The previous steps allows us to estimate θ from the differential equation (9) ; we denote this series of estimators $\hat{\theta}^{(2)}$. γ is now estimated by the following minimization procedure:

$$\hat{\gamma} = \arg\min_{\gamma \in [0,1[} \sum_{i=1}^n \left(\hat{\theta}_i^{(2)} - \hat{\theta}(\gamma)_i\right).$$

5.4 A short Monte Carlo exercise

The results are based on $n_s = 100$ replications of samples of size $n = 2000$. In this set-up the marginal cost function is a linear function $C'_{0q}(q, W) = q + W$, where $W \sim \mathbb{U}_{[0,1]}$. θ the type of the firm is such that $\theta \sim B(2, 2)$ with a translation to ensure that the median is equal to 1. $S'_0(q, Z) = 15 - 5q + Z$, where $Z \sim N(0, 1)$ and $\gamma = 0.2$.

Table displays the integrated square bias, integrated variance and integrated mean square error based on 100 simulations for the sieve IV estimator of the marginal cost function $C'_{0q}(\cdot)$. Different values for K_n, J_n and κ are presented to control for the influence of each parameter. It appears that the results ......

	Integrated MSE					
n_m		deg=9			deg=16	
2	0.9166	0.8803	0.8683	0.4423	0.4175	0.4101
5	0.9146	0.8788	0.8675	0.4420	0.4179	0.4103
8	0.9157	0.8767	0.8648	0.4431	0.4183	0.4106
13	0.9176	0.8788	0.8638	0.4386	0.4205	0.4117

Env:6.000947

6   Empirical Application

We apply the estimation method detailed above to the case of water utilities in France. We first present the data set, before illustrating the application of our econometric method to the estimation of the structural model of regulation.

6.1   The Data

As already mentioned, the production and distribution of water to households in France are decided at a local level. A survey has been conducted by the French environment institute (IFEN, Institut Français de l'Environnement) amongst local municipalities in metropolitan France.[23] All municipalities with a population of more than 10 000 are in the survey and the sampling rate is decreasing with the population. It is given in Appendix **??**. The sample is therefore representative from the population of French local communities (36 203 local municipalities in 2001).

We select the observations for which the chosen mode is either operating through a public company (*"régie"*) or through a private operator with a lease contract (this is the majority of the cases). Then we select the typical contracts, i.e. the two-part tariff contracts (around 95% of the observations). We also merge our dataset with an administrative dataset which reports the median income in a given municipality. For anonymity reasons, it is not reported for city size less than 150 inhabitants. Our sample does not contain such small villages. In the end, we have 3687 observations which represent around 27000 local municipalities. Table 1 presents the management mode of municipalities ranked according to their sampling rate.

In our dataset, we observe the fix fee of the contract A and the variable part p paid by meter-cube consumed. The overall quantity q is also observed. The characteristics of the network are also given (length and density, a measure of the age of the infrastructure, % of

---

[23]excluding Corsica and overseas departments are excluded from our study.

Table 1: Management mode by municipality sizes

	Population N of the different municipalities					
	$N < 400$	$400 \le N$ $N < 1\,000$	$1\,000 \le N$ $N < 2\,000$	$2\,000 \le N$ $N < 3\,500$	$3\,500 \le N$ $N < 10\,000$	$10\,000 \le N$
Sampling Rate	4.80	8.37	14.64	22.72	88.03	100.00
Number n of municipalities						
n	581.00	663.00	489.00	348.00	1130.00	476.00
Eq. number	12105.00	7926.00	3340.00	1532.00	1284.00	476.00
Public	51.12	48.11	43.56	35.63	33.63	34.03
Delegation	48.88	51.89	56.44	64.37	66.37	65.97
Providers in the delegation mode						
Veolia	35.56	31.69	39.13	41.07	45.87	55.41
Saur	37.68	34.30	29.35	29.02	17.47	10.51
Suez	14.79	20.35	22.83	20.98	25.73	23.57
Other	11.97	13.66	8.70	8.93	10.93	10.51

leaks, quality of the water used as an input, treatment applied before distribution). Some additional information related to the local community (population, median income, size of the city) is recorded. We present in Table 2 some descriptive statistics of these variables. For continuous variables, we present the median and the interquartile range $Q_3 - Q_1$ which are respectively robust measures of the location and the dispersion of the empirical distribution. For the two qualitative variables related to the quality of the water and the treatment (they are both recoded with three modalities), we report the empirical frequencies.

The fix fee represents around one quarter of the total transfer paid to the provided. The variable part is around 1 euro per m^3. Price paid by the consumer is generally lower in a local municipality which is publicly managed. On the other hand, private firms has a higher probability to operate in more dense cities and to distribute water which requires heavier treatment to be drinkable.

6.2   Estimation of the model

We now follow the estimation strategy that is exposed in Section 5. The 95% confidence intervals are computed using the bootstrap procedure of Chen and Pouzo (2013).

This section is currently under investigation.

Table 2: Descriptive statistics of the main variables

	All mode		Public		Delegation	
	Median	$Q_3 - Q_1$	Median	$Q_3 - Q_1$	Median	$Q_3 - Q_1$
A	35	31.50	27	28	40	37
p	0.94	0.41	0.82	0.36	1.02	0.41
$A + 120p$	150	62.30	134.40	48.90	166.80	64.20
q	128.57	56.34	126.21	54.84	130.16	57.08
Pop	2384	5274	1523	4537	3307	5742.50
Med. Income	22479	7655.50	22276	6949.50	22609.50	8206.50
Network characteristics						
Area	1428	1675.50	1436	1616.50	1420	1733.25
Network length	34	52	28	47	39	55
Vetuste	30.00	70.00	50.00	65.00	15.00	65.00
	Frequency		Frequency		Frequency	
Deep water	71.47		73.78		69.89	
Sup. water	10.63		9.36		11.50	
Basic Tt	53.27		52.04		54.11	
High Tt	17.71		10.17		22.86	

7 Counterfactuals

.

8 Alternative Formulations

This section discusses alternative formulation for the contracting environment and how our identification procedure applies or not. While Section **??** highlights environment where our basic identification strategy would still be useful, Section 8.3 shows the importance of the kind of contracts observed (relying on fixed fees and per-unit of consumption prices) to get such positive results.

8.1 Non-Separability in the Cost Function

Let us now suppose that the cost function is no longer separable and can be more generally written as $C(q, \theta)$, i.e., costs are not necessarily linear in θ. We assume that $C(q, \theta)$ is

increasing and concave in q with on top $C'_\theta > 0$ (operators with lower types produce at lower costs) and the Spence-Mirrlees condition $C''_{q\theta} > 0$ (those operators also produce at lower marginal costs) being satisfied. It is routine to check tha the system (9)-(11)-(12) now becomes:

$$A'_p(p(\theta, \varepsilon), \varepsilon) = -\left(p(\theta, \varepsilon) - C'_q(D(p(\theta, \varepsilon), \varepsilon), \theta)\right) D'_p(p(\theta, \varepsilon), \varepsilon) - D(p(\theta, \varepsilon), \varepsilon). \quad (25)$$

$$p(\theta, \varepsilon) = C'_q(D(p(\theta, \varepsilon), \varepsilon), \theta) + (1 - \gamma)\frac{F(\theta)}{f(\theta)}C''_{q\theta}((D(p(\theta, \varepsilon), \varepsilon)), \theta), \quad (26)$$

$$S'_q(D(p(\theta, \varepsilon), \varepsilon), \varepsilon) = p(\theta, \varepsilon), \quad (27)$$

8.2   Identification of a More General Specification

We now prove the identification of this more general model with non-separability in both the surplus and cost functions (Equations (25)-(26)-(27)). First, observe that the identification of the marginal surplus is similar in this general model to what we have done above.

For the marginal cost function, we need an additional normalization like for the surplus function. We shall assume that θ is assumed to be uniformly distributed on $[0, 1]$ and we therefore identify the marginal cost function (given W) for the α-th quantile firm in an efficiency ranking (ranked from the most efficient , $\alpha = 0$, to the less efficient, $\alpha = 1$).

Equation (25) identifies $c'_q(q, \alpha, W)$ for any quantile α of θ under an assumption of completeness of $\tilde{Z}$ in q. We can indeed derive a moment condition similarly than what have been done in the seprable case earlier. We are indeed able to know, after having derived the marginal surplus function for any quantile of ε the quantity $d = p(\theta, \varepsilon) - \frac{A'_p(p(\theta, \varepsilon), \varepsilon) - D(p(\theta, \varepsilon), \varepsilon)}{D'_p(p(\theta, \varepsilon), \varepsilon)}$, which is equal to the marginal cost (given W) for some (unknown) quantile of θ. Moreover the monotonicity of the marginal cost in θ (given W) ensures that (see above for a similar derivation):

$$\mathbb{P}\left(\left(c'_q(\alpha, q, W) \leq d\right) | Z, W\right) = \alpha. \quad (28)$$

The completeness assumption ensures therefore the identification of any marginal cost function.

Coming back now to Equation (26), the only unknown quantity is now γ and we also recover the identification of the bargaining power in the non-separable case.

8.3 Contracting on Outputs

The fact that the model can be fully identified (at least when $\gamma = 0$) might seem quite surprising at first glance. Indeed, the literature on the structural estimations of incentive contracts has repeatedly failed to obtain such joint identification of cost functions and types distributions because the data available to the econometrician do not provide enough information. The key difference between our model and those of the existing literature, especially d'Haultfoeuille and Février (2010) and Vuong and Perrigne (2012), comes actually from the set of contracting variables available under different scenarios that fit with specific institutional environments.

In the regulatory context under scrutiny, for instance, consumers adopt a competitive behavior, expressing individual demand for water at the stipulated unit price while the operator stands ready to supply aggregate demand at that a market clearing price. Instead, had the principal being a single big customer dealing directly with the operator, contracting on the quantity bought would be a feasible option. This is such scenario that is analyzed in d'Haultfoeuille and Février (2010) and Vuong and Perrigne (2012).

To facilitate comparison with Section 3.1, let us thus assume that contracts are based only on the operator's quantity. In this hypothetical scenario where output could be directly contracted upon, a nonlinear contract might now specify a payment $T(q)$ to the operator if he offers a volume q.[24] For the sake of simplifying exposition, we also assume in this section that there is no fluctuations in demand (i.e., $\varepsilon \equiv 0$) and $\gamma = 0$. Under those conditions, we already know from Section 4.1 that our base model is actually identified.

Theoretical results. Mimicking some of the earlier steps of our above analysis, we can rewrite the operator's information rent and optimal supply respectively as:

$$U(\theta) = \max_q T(q) - \theta C_0(q) \tag{29}$$

and

$$q(\theta) = \arg\max_q T(q) - \theta C_0(q). \tag{30}$$

From there, it again follows from incentive compatibility that $U(\cdot)$ is absolutely continuous

---

[24]Observe that, even in this output scenario, the principal still does not observe costs. Otherwise, he could retrieve from the joint observation of costs and outputs information on the agent's type. See Laffont and Tirole (1986) for a theoretical model where such cost observation is available and yet, because of an extra moral hazard variable, asymmetric information still matters. Perrigne and Vuong (2011) for the corresponding empirical study.

and admits an integral representation as:

$$U(\theta) \int_{\theta}^{\bar{\theta}} C_0\left(q(x)\right) dx \tag{31}$$

where again, we take into account that the operator's participation constraint is binding at $\bar{\theta}$ for the optimal contract. Observe that this rent is again greater as the operator is requested to produced more.

Under asymmetric information, an optimal contract maximizes the expected welfare of the municipality subject to incentive and participation constraints. From our above observations, that incentive feasible set can be summarized by constraints (31) and a monotonicity condition (namely $q(\theta)$ weakly decreasing in θ) that will be omitted in a first step and checked ex post on the solution to the solution of the so relaxed problem. Formally, this relaxed problem can be written as:

$$\max_{\{q(\cdot),U(\cdot)\}} \int_{\underline{\theta}}^{\overline{\theta}} \left[S(q(\theta)) - \theta C_0\left(q(\theta)\right) - U(\theta)\right] dF(\theta) \text{ subject to (31)}.$$

Using (31), integrating by parts as we did above and optimizing pointwise leads to the following expression of the optimal output $q(\theta)$:

$$S'_q(q(\theta)) = \left(\theta + \frac{F(\theta)}{f(\theta)}\right) C'_{0q}\left(q(\theta)\right). \tag{32}$$

The structural model is then defined also by appending to that optimality condition for the principal's problem a second optimality condition related to the agent's problem, namely:

$$T'_q(q(\theta)) = \theta C'_{0q}\left(q(\theta)\right). \tag{33}$$

From an economic viewpoint, whether outputs as here or prices as in our base line model are the relevant contracting variables leads to the same optimal outputs. The principal's optimality condition (32) is similar to that obtained in our base model (11) (for $\gamma = 0$) and downward output distortions also follow from replacing cost parameter by their virtual counterparts. Since optimal outputs in the two models are the same, the agent's information rents, consumer's net surpluses and welfares are also identical. This economic equivalence should come at no surprise. Whether the demand side of the market adopts a "command and control" approach to specify how much should be supplied to satisfy aggregate needs or whether individual consumers are left to express demand is equivalent since that demand side is not the source of any informational asymmetry.

28

Differences in the scope for identification. Although economically equivalent, the two contracting scenarios differ not only in terms of the set of observables that are available to the econometrician but also in terms of the possibilities left for identification. In other words, whether prices or quantities are controlled gives two alternative implementations of the same allocation which are equivalent from the point of view of the players in the regulatory game which are equally informed on cost and surplus functions and distribution functions.

Yet, those two implementations differ from the perspective of the econometrician. The econometrician must infer from market conduct some information that is common knowledge among players. At a rough level, this outside observer faces a much harder signal extracting problem. In the scenario where quantities are directly controlled, no information on marginal surplus can be learned from the firm's choices. Instead, when only prices are controlled, information on marginal surplus gets communicated to outside observer. This simple fact allows identification of the model.

More precisely, for the output scenario, the only observables are the output q and the overall payment $T(q)$. As in our base model, S_q' and C_{0q}' are not observed in the data so that the principal's optimality condition (32) cannot be used to retrieve information either on types or on their distribution. On top, per-unit consumption prices are by construction not observed, a key difference with our base model.

Defining an implicit per-unit consumption price for this output scenario as $p(\theta) = S_q'(q(\theta))$, we may follow a procedure similar to that developed in our base model and introduce a fictitious price-cost margin. Accordingly, we thus rewrite the principal's optimality condition (32) as a familiar condition:

$$\frac{p(\theta) - \theta C_{0q}'(q(\theta))}{p(\theta)} = r(\theta, 0).$$

Even if such price is not observed, one could hope as in our base line model to retrieve the relevant information from the agent's optimality condition (33). However, this step is no longer possible here because the agent's optimality condition (33) now refers only to the marginal nonlinear price $T_q'(q(\theta))$ and not on the price as in our baseline scenario. Indeed, this marginal nonlinear price $T_q'(q(\theta))$ always differs from the per-unit consumption price $p(\theta) = S_q'(q(\theta))$ under asymmetric information, there is always a wedge between $T_q'(q(\theta))$ and $p(\theta) = S_q'(q(\theta))$, as:

$$p(\theta) - T_q'(q(\theta)) = \frac{F(\theta)}{f(\theta)} C_{0q}'(q(\theta)) > 0.$$

Thus, the agent's optimality condition does not bring new information on marginal surplus at the equilibrium point, making it impossible to identify this function.

In response to such impossibility, different alleys have been pursued in the literature. Vuong and Perrigne (2011) consider a model of "false moral hazard"[25] where costs, which results from innate types but also no verifiable effort variables, can be observed. Such observation of course brings extra information. Although positive results are now obtained, the analysis remains complex because the additional effort variable that characterizes such model must also be disentangled from innate types. Developing an approach tailored to the specificities of their data set, d'Haultfoeuille and Février (2010) work with some discrete heterogeneity on the demand side while still assuming that such information is observed by the econometrician.[26] Three points are in this case sufficient to recover identification. Finally, Vuong and Perrigne (2012) consider a nonlinear pricing model with fixed quantity which *de facto* determines marginal surplus at the equilibrium point. Let us conclude this section by pointing out an alternative route. Indeed, parametric assumptions on both the surplus function and the cost function, or the cost function and the firms' types could help identifying the model but, of course, at a risk of misspecification as these assumptions cannot be tested.

9 Conclusion

To be continued...

---

[25]In the parlance of the incentive literature, see Laffont and Martimort (2002, Chapter 7).

[26]This is in striking contrast with our assumption in the base line model that such heterogeneity is actually not observed.

References

[1] Ai, and X. Chen (2003), "Moment Restrictions Containing Unknown Functions," *Econometrica*, 71: 1795-1843.

[2] Armstrong, M. and D. Sappington (2007), "Recent Developments in the Theory of Regulation," *Handbook of Industrial Organization, Vol. 3*, M. Armstrong and R. Porter eds. North-Holland.

[3] agnoli, C. and T. Bergstrom (2005), "Log-Concave Probability and its Applications," *Economic Theory*, 26: 445-469.

[4] Baron, D. (1989a), "Design of Regulatory Mechanisms and Institutions" in *Handbook of Industrial Organization*, vol. 2, 1347-1447.

[5] Baron, D. (1989b), "Regulation and Legislative Choice," *The RAND Journal of Economics,* 19: 467-477.

[6] Baron, D. and R. Myerson (1982), "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50: 911-930.

[7] Basov, S. (2005), *Multidimensional Screening,* Springer-Verlag.

[8] Bhattacharyya, A., T.R. Harris, R. Narayan and K. Raffiee (1995), "Specification and Estimation of the Effect of Ownership on the Economic Efficiency of the Water Utilities," *Regional Science and Urban Economics*, 25: 759-784.

[9] Bhattacharyya, A., E. Parker and K. Raffiee (1994), *An Examination of the Effect of Ownership on the Relative Efficiency of Public and Private Water Utilities*, Land Economics, 70; 197-209.

[10] Blundell, R., Chen, X. and D. Kristensen (2007), "Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves," *Econometrica*, 75: 1613-1669.

[11] Brocas, I., K. Chan and I. Perrigne (2006), "Regulation under Asymmetric Information in Water Utilities," *American Economic Review,* 96: 62-66.

[12] Carrasco, M., Florens, J.-P. and E. Renault (2008), "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics,* Vol 6..

[13] Cazals, C., J.P. Florens and L. Simar (2002), "Nonparametric Frontier Estimation: a Robust Approach," *Journal of Econometrics*, 106: 1-25.

[14] Chen, X. and D. Pouzo (2008), "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Non-Smooth Moments," Cowles Foundation Discussion Paper 1640.

[15] Chernozhukov, V., P. Gagliardiniy and O. Scailletz (2008), "Nonparametric Instrumental Variable Estimation of Quantile Structural Effects," Working paper.

[16] Chernozhukov, V. and C. Hansen (2005), "An IVModel of Quantile Treatment Effect," *Econometrica,* 73: 245-261.

[17] Chernozhukov, V., Imbens, G. and W. Newey (2007), "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics,* 139: 4-14.

[18] Chesher, A. (2003), "Identification in NonSeparable Models", Econometrica, 71: 1405-1441.

[19] Chesher, A., (2007), "Identification of Nonadditive Structural Functions," in *Advances in Economics and Econometrics, Theory and Applications: 9th World Congress of the Econometric Society,* Blundell, R., Persson, T. and W. Newey, Eds., Vol. 3, Cambridge University Press, Cambridge.

[20] Chong, E., F. Huet, S. Saussier and F. Steiner (2006), "Public-Private Partnerships and Prices: Evidence from Water Distribution in France," *Review of Industrial Organization* 29: 149-169.

[21] Commissariat Général au Developpement Durable (2010), *Le point Sur*, Vol. 67.

[22] Darolles, S., Florens, J.-P. and E. Renault (2003) "Nonparametric Instrumental Regression, " Working Paper.

[23] Estache, A. and M. Rossi (1999), "Comparing the Performance of Public and Private Water Companies in Asia and Pacific Region: What a Stochastic Costs fFontier Shows," *World Bank working paper* WPS 2152, Washington, DC.

[24] Gagnepain, P. and M. Ivaldi (2002), "Incentive Regulatory Policies: The Case of Public Transit Systems in France," *The RAND Journal of Economics,* 33: 605-629.

[25] Gagnepain, P., M. Ivaldi and D. Martimort (2013), " The Cost of Contract Renegotiation: Evidence from the Local Public Sector," forthcoming *The American Economic Review.*

[26] Garcia, S. and A. Thomas (2001), "The Structure of Municipal Water Supply Costs: Application to a Panel of French Local Communities," *Journal of Productivity Analysis*, 16: 5-29.

[27] Hall, D. (2000), "Public Choice and Water Rate Design, " in *The Political Economy of Water Pricing Reforms*, Ariel Dinar (ed.): 189-212. Oxford University Press, Oxford.

[28] d'Haultfoeuille X. and P. Fevrier (2007), "Identification and Estimation of Incentive Problems: Adverse Selection," Working Paper CREST Paris.

[29] IFEN (Institut Français de l'Environnement) (2006), "Les Services Publics de l'Eau en 2004 - Volet Eau Potable," Paris.

[30] Imbens, G. and W. Newey (2002), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper.

[31] Koenker, R. (2005), "Quantile Regression," Econometric Society Monographs, 38, Cambridge University Press, Cambridge.

[32] Koenker, R., Ng., P. and S. Portnoy (1994), "Quantiles Smoothing Splines," *Biometrika* 81: 673-680.

[33] Laffont, J.J. and D. Martimort (2002), *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.

[34] Laffont, J.J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.

[35] Lavergne P. and A. Thomas (2005), "Semiparametric Estimation and Testing in Models of Adverse Selection," *Empirical Economics,* 30: 1-22

[36] Myerson, R. (1982), "Optimal Coordination Mechanisms in Generalized Principal-Agent Models,"*Journal of Mathematical Economics,* 10: 67-81.

[37] Noldeke, G., and L. Samuelson, (2007), "Optimal Bunching Without Optimal Control," *Journal of Economic Theory*, 134: 405-420.

[38] Perrigne, I. and Q. Vuong (2012), "Nonlinear Pricing in Yellow Pages," Working paper, Pennsylvania State University.

[39] Perrigne, I., and Q. Vuong (2011), "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard," *Econometrica*, 79: 1499-1539.

[40] Robinson, P. (1988), "Root-n Consistent Semiparametric Regression," *Econometrica*, 56: 931-954.

[41] Wolak, F. (1994), "An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction," *Annales d'Economie et de Statistique,* 34: 13-69.

Table 3: Average price, volume and community characteristics, by management mode

Variable	Management mode	
	Public	Private
Median income	23,177.86	24,013.1
Population	4592.556	5450.189
Customers	1900.296	2187.716
Tariff fixed part (Euro/year)	32.2135	47.4118
Tariff variable part (Euro/m^3)	1.0801	1.3578
Average consumption (m^3/customer/year)	136.8093	141.2813
Average price (Euro/m^3)	1.3562	1.7448
Customer network density (Customers/network length)	39.0283	38.6851

Notes. Source: IFEN (2006). Tariff components and average prices include tax but concern only water supply (and not sanitation).

Table 4: Polling rates and corresponding weights

N		400	1000	2000	3500	
	N<	≤ N<	≤N<	≤N<	≤N<	N>
	400	1000	2000	3500	10,000	10,000
(number of inhabitants)						
Sampling rate (percent)	4.8	8.4	14.6	22.7	88.0	100
Weight	20.83	11.95	6.83	4.40	1.13	1