Identification and Estimation of Incentive Problems: Adverse Selection[*]

Xavier d'Haultfoeuille [†]   Philippe Février[‡]

September 2007

Abstract

The adverse selection model is a principal-agent model defined by the objective function of the principal, the agents' utility function and the distribution of agents' types. We prove that the nonparametric identification of this model requires the knowledge of at least one of the three functions. We also show that some exogenous changes in the objective function of the principal are sufficient to obtain partial or full nonparametric identification of the model. A nonparametric estimation procedure based on these results is proposed. We apply this method to test if firms provide the right incentives to their workers. Using contract data between the French National Institute of Statistics and its interviewers, we test and reject the contracts' optimality. We estimate, however, the loss of using linear contracts instead of optimal ones to be less than 10%, which may explain why these simple contracts are so popular.

Keywords: Adverse Selection, Nonparametric Identification, Nonparametric Estimation, Asymmetric Information, Incentives, Optimal Contracts.

JEL classification numbers: C14, D82, D86

1   Introduction

Since the seminal work of Akerlof (1970), extensive attention has been devoted to asymmetries of information and their consequences in economics. A canonical example where these asymmetries play a fundamental role is the adverse selection model. This model is helpful, for instance, to better understand nonlinear pricing, regulation, financial contracts or taxation theory (Wilson, 1993; Laffont and Tirole, 1993; Freixas and Rochet, 1997; Diamond, 1998). However, and as pointed out by Chiappori and Salanié (2002), few econometric work has been done to estimate structurally these models. Regulatory contracts have been studied by Wolak (1994), Gagnepain and Ivaldi (2002) and Perrigne (2002), while Ivaldi and Martimort (1994) and Miravette (2002) estimated nonlinear pricing models. All these papers adopt a parametric framework. Lavergne and Thomas (2005) are more flexible and specify a semiparametric model to study regulation. Perrigne and Vuong (2004) are the only ones who follow a nonparametric approach. They study the Laffont and Tirole (1986) regulation model in which ex-post costs are observed, and show that such a model is nonparametrically identified.

In this paper, we study the empirical content of the canonical adverse selection model (Laffont and Martimort (2002), Salanie (2005a)). This model is characterized by the objective function of the principal, the distribution of agents' types and the utility function of the agents. The econometrician is supposed to observe the contract and the associated trades. In the most general setting, the model is not identified. However, under a separability assumption on the utility of the agents, we prove that the knowledge of one of the structural functions is sufficient to obtain full identification. Hence, in a regulation context for instance, if ex-post costs are observed, the utility function can be recovered and full identification is achieved. Similarly, if the regulatory maximizes the sum of the firms' profits and of the consumers' surplus, the objective function of the principal is known and the model is nonparametrically identified.

The identification of the model can also be achieved by observing a change in the contracts between the principal and the agents, under the exclusion restriction that the utility function and the types of the agents are not affected by this change. As described by Chiappori and Salanié (2002) such exclusion restrictions arise naturally in experiments or natural experiments (see e.g. Manning et al., 1987, Ausubel, 1999, Lazear, 2000, or Shearer, 2004). We characterize what is identified under these conditions. If the marginal transfer functions defined in the two contracts cross, the model is fully identified. If not, nonparametric bounds can be recovered on the utility function and the distribution of the types. Fur-

thermore, two changes can be sufficient to obtain full nonparametric identification of the model.

To prove these results, we extensively use the first order condition which defines the optimal choices of the agents, and the observed distribution of the trades. The first equation allows us to define what we call horizontal transformations whereas the second one yields to vertical transformations. These transformations are identified in the data and are combined to define recursively the functions of interest. An important feature of our identification procedure is that the utility function and the distribution of the agents' types are recovered using the agent's program solely. This is convenient when the optimality of the principal is questionable. For instance, the common knowledge assumption on the distribution function of the agents' types or their cost function may fail to hold, the principal may also be risk averse (see Lewis and Sappington, 1995, Gence-Creux, 2000) and the costs of implementing nonlinear contracts may modify significantly his program (see Ferrall and Shearer, 1999). Our results are not affected by these problems.

Beyond identification, we also examine some tests of the model. First, our identification results can be used to test in a structural and nonparametric way the optimality of the observed contracts. This question has received considerable attention in the empirical literature (see Prendergast, 1999, for a survey) and the results are rather mixed, the authors finding few evidence that contracts vary with the relevant parameters. However, no nonparametric structural test of the model has been proposed so far.[1] Second, we examine the possibility of testing the asymmetric model versus the symmetric one. We show that implementing such a test is difficult. Without prior knowledge on the structural functions, both models cannot be distinguished nonparametrically, even with exogenous variations in the contracts. This result contrats with previous papers on the topic (see e. g. Puelz and Snow, 1994, Chiappori and Salanié, 2000) in which auxiliary variables are used.

We also develop a nonparametric estimation procedure based on our recursive identification method and prove that the estimators are consistent. This method is implemented on contract data between the French National Institute of Economics and Statistics (Insee) and its interviewers[2] to study incentives in firms in the spirit of Ferrall and Shearer (1999) and Paarsch and Shearer (2000). Thanks to a change in the bonus paid to interviewers, we recover bounds on the utility function of the agents and the distribution of their type. Then, using the objective function of the principal, we are able to perform a nonparametric test of the contracts' optimality. We reject that Insee's linear contracts are optimal

---

[1]Ferrall and Shearer (1999) implement a structural but parametric test.

[2]Insee hires interviewers for its household surveys.

and concludes that it does not fully take into account agents' responses when writing his contracts. However, when estimating the cost of using linear contracts instead of the optimal ones, we find that Insee's loss is about 9%. This result contrasts with the previous literature in which linear contracts were thought to be quite inefficient but simple and easy to implement (Ferrall and Shearer, 1999). Our application points out on the contrary that the loss is quite small, which may explain the wide use of linear contracts. We also recover what Insee's surplus would have been under complete information and find that the estimated expected surplus under incomplete information are 84% of full information surplus. Overall, the cost due to the asymmetry of information are twice the costs associated with the bonus system. Our last results are obtained using parametric specifications in line with our nonparametric results.

The paper is organized as follows. Section 2 recalls the main theoretical results for a principal-agent model with adverse selection. Section 3 is devoted to the nonparametric identification of this model. Our nonparametric estimation method and its application are presented in section 4, and section 5 concludes.

2 Adverse selection model

Following Laffont and Martimort (2002) and Salanié (2005a), we consider a basic adverse selection model where a principal trades y with some agents and provides them with a monetary transfer t. Agents are heterogeneous with a quasi-linear utility function $U(t, y, \theta) = t - C(y, \theta)$.[3] The monetary cost $C(y, \theta)$ of implementing y depends on their type θ which is unobserved by the principal. We suppose that θ is real and nonnegative, $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, so that we can interpret it as a measure of the agent's intrinsic efficiency. $\underline{\theta}$ is the most efficient agent's type whereas $\overline{\theta}$ is the least efficient. We denote by $F_\theta(.)$ (resp. $f_\theta(.)$) the distribution function (resp. density function) of θ and suppose it to be common knowledge. Lastly, the principal is assumed to be risk neutral. His objective function is quasi linear and takes the form $W(t, y, \theta) = S(y, \theta) - t$. The following regularity conditions are imposed.

Assumption 1 *(regularity conditions)* $f_\theta(.) > 0$; $\partial S/\partial y(.,.) > 0$ and $\partial^2 S/\partial y^2(.,.) < 0$; $\partial C/\partial \theta(.,.) > 0$, $\partial C/\partial y(.,.) > 0$, $\partial^2 C/\partial y^2(.,.) > 0$ and $\partial^2 C/\partial \theta \partial y(.,.) > 0$.

---

[3]The convention, here, is that y is produced by the agents as in the regulatory model. Equivalently, we could assume that the agents consume y and that the utility function takes the form $U(t, y, \theta) = U(y, \theta) - t$ as in the price discrimination model.

The objective function of the principal is increasing and concave. The cost increases with inefficiency and with the level of y. Moreover, it is convex as a function of y. Lastly, the positivity of its cross derivative is the Spence-Mirrlees condition, which indicates that a more efficient type is also more efficient at the margin.

The firm proposes to the agent a set of contracts of the form $[(y, t); y \in \mathbb{R}^+, t \in \mathbb{R}^+]$. The agent of type θ can either refuse all contracts or accept one of them. If he accepts a contract (y, t), the agent delivers y and receives a transfer t. If he refuses, he obtains his outside opportunity utility level normalized to zero.

Without asymmetry of information between the principal and the agent, the firm makes a take it or leave it offer to the agent of type θ that implements first-best trade levels. More precisely, the firm proposes to an agent of type θ a contract (y_θ^*, t_θ^*) defined by :

$$\frac{\partial S}{\partial y}(y_\theta^*, \theta) = \frac{\partial C}{\partial y}(y_\theta^*, \theta) \tag{2.1}$$

$$t_\theta^* = C(y_\theta^*, \theta) \tag{2.2}$$

The optimal trade level is the quantity that equalizes marginal gain and marginal cost. The transfer function is such that the agent accepts the offer but makes zero profit.

If θ is the agent's private information, the complete information optimal contracts can no longer be implemented. The problem arises because of the asymmetric information. In particular, efficient agents mimic inefficient ones and prefer to trade less to have a positive utility.

The optimal menu of contracts is more complex in this case. It may be the case, for example, that the contracts targeted for different types coincide. For those contracts, we say that there is bunching of types. The general theory can be found in Laffont and Martimort (2002) but we only describe the optimal menu of contracts without bunching.

Proposition 2.1 *Laffont-Martimort (2002)*
Under assumption 1 and if there is no bunching at equilibrium, the optimal menu of contracts, of the form $(y(\theta), t(\theta); \theta \in \Theta)$, entails:

- *A downward output distortion for all types except the most efficient:*

$$\frac{\partial S}{\partial y}(y(\theta), \theta) = \frac{\partial C}{\partial y}(y(\theta), \theta) + \frac{F_\theta(\theta)}{f_\theta(\theta)} \frac{\partial^2 C}{\partial \theta \partial y}(y(\theta), \theta) \tag{2.3}$$

- *A positive information rent for all types except the less efficient:*

5

$$t(\theta) = C(y(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \frac{\partial C}{\partial \theta}(y(\tau), \tau)d\tau \qquad (2.4)$$

The firm has to leave a positive information rent to the agents for them to reveal their types (the term $\int_{\theta}^{\bar{\theta}} \partial C/\partial \theta (y(\tau), \tau)d\tau$ in equation (2.4)). This information rent increases with the efficiency of the agent and creates inefficiencies in production (the term $F_{\theta}(\theta)/f_{\theta}(\theta) \times \partial^2 C/\partial \theta \partial y(y(\theta), \theta)$ in equation (2.3)).

3 Nonparametric identification

3.1 The general case

We now discuss the empirical content of the model. In the sequel, we suppose that the econometrician observes the trades at equilibrium $(y(\theta_i))_{i \in \mathbb{N}}$ for an infinite sample of agents indexed by i (the types θ_i being independently drawn from F_{θ}). Agents are supposed to be homogenous except for their unknown types; if they differ by observed characteristics, our results below must be understood conditionally on these characteristics. We also suppose that the corresponding transfers $t(y(\theta_i))$ are observable.[4] The trades and transfers enable one to identify the cumulative distribution function of y, $F_y(.)$ and the transfer function $t(.)$ on the support $\mathcal{Y}$ of $y(.)$. The question is whether $C(.,.)$, $F_{\theta}(.)$ and $S(.,.)$ can be recovered from these functions and the model.

Without further assumption, we can always replace θ by $F_{\theta}(\theta)$ and change $C(.,.)$, F_{θ} and $S(.,.)$ accordingly. Consequently, F_{θ} is not identified and we can suppose θ to be uniformly distributed on $[0, 1]$. Hence, we focus on $C(.,.)$ and $S(.,.)$ only.

Furthermore, we analyse in this section the case where there is no bunching at equilibrium. Under assumption 1, the existence of bunching is equivalent to F_y admitting at least one mass point. Hence, it is easily testable in the data, by checking whether at least two observed trades are identical or not. The analysis of bunching is deferred to subsection 3.4.

Assumption 2 *There is no bunching at the equilibrium.*

Our identification results are based on three equations.

---

[4]This assumption may be strong (see Wolak, 1994, and Ferrall and Shearer, 1999, for examples where the transfers are unknown).

First, by the Spence-Mirrlees condition and assumption 2, $y(.)$ is strictly decreasing. Thus it admits an inverse $\theta(.)$ which satisfies, for all $y \in \mathcal{Y}$,

$$F_y(y) = \mathbb{P}(y(\theta) < y) = \mathbb{P}(\theta > \theta(y)) = 1 - \theta(y) \tag{3.1}$$

The first equality stems from the fact that the distribution of y is atomless and the second from $\theta(.)$ being strictly decreasing. Because $F_y(.)$ is observed, this equation shows that $\theta(.)$ is identified .

Secondly, the agent chooses his production according to the first order condition, which writes as

$$\frac{\partial C}{\partial y}(y, \theta(y)) = t'(y) \tag{3.2}$$

In other terms, $\partial C/\partial y$ is identified on $L = \{(y, \theta(y)), y \in \mathcal{Y}\}$. Note also that $\partial C/\partial y$ is not identified elsewhere because no (y, θ) is observed outside equilibrium. Given the price schedule, we can only recover for each agent's type θ, the marginal utility at his optimal choice of production.

Lastly, the third equation is the first order condition of the principal (2.3), which can also be written as:

$$\widetilde{S}(y) = \frac{\partial C}{\partial y}(y, \theta(y)) + \theta(y)\frac{\partial^2 C}{\partial\theta\partial y}(y, \theta(y)) \tag{3.3}$$

where $\widetilde{S}(y) = \partial S/\partial y(y, \theta(y))$. From an identification point of view, this equation does not impose conditions on $S(.,.)$ directly, but rather on $\widetilde{S}(.)$, and we thus focus on this function hereafter.[5] Because no (y, θ) is observed outside equilibrium, $\partial^2 C/\partial\theta\partial y$ is not identified. Thus, $\widetilde{S}(.)$ is not identified either.

Note that even if $\widetilde{S}(.)$ is known by the theory, we cannot recover $\partial C/\partial y(.,.)$ outside L and predict for instance what would be the optimal contract under another objective function of the principal.

To circumvent this nonidentification result, we impose a restriction on the utility function by supposing that the cost function is separable. This hypothesis is often made in the theoretical adverse selection literature (see Wilson, 1993 or Laffont and Tirole, 1993). It is also assumed in empirical research (see Wolak, 1994, Ferrall and Shearer, 1999, Lavergne and Thomas, 2005) and in the nonparametric analysis of Perrigne and Vuong (2004). Other functional restrictions on $C(.,.)$ are possible. We present one of them, which stems from the false moral hazard model, in subsection 3.4.

---

[5]In general, the knowledge of $\widetilde{S}(.)$ does not enable to recover $S(.,.)$. However, it will (up to an additive constant) in some applications where $S(.,.)$ does not actually depend on θ.

Assumption 3 *(cost separability)* $C(y, \theta) = \theta C(y)$.

Under assumption 3, the uniform normalization of F_θ is not possible anymore. On the other hand, we can replace $(\theta, C(.))$ by $(\alpha\theta, C(.)/\alpha)$ and leave the model unchanged. Thus, another normalization is necessary and for a given $y_0 \in \mathcal{Y}$, we can choose θ_0 such that $\theta(y_0) = \theta_0$. In other words, assumption 3 reduces the dimensionality of the problem by replacing a function of two variables $C(.,.)$ by two functions of one variable $C(.)$ and $F_\theta(.)$ and the structural parameters are now $(C', F_\theta, \widetilde{S})$. Identification is based on the same equations than previously, and the proof of proposition 3.1 is deferred to appendix A.

Proposition 3.1 *Under assumptions 1, 2 and 3, $(C', F_\theta, \widetilde{S})$ are not identified jointly. On the other hand, if one of these three functions is known, the other two can be identified.*

This result states that even under the separability assumption, the model remains unidentified. Basically, only three equations are available whereas we seek to recover four functions: $(C'(.), F_\theta(.), \widetilde{S}(.), \theta(.))$. Hence, we can fix one of the four functions and deduce the others from the data and the model. Actually, this function cannot be chosen completely arbitrarily. Indeed, the three structural functions $(C'(.), F_\theta(.), \widetilde{S}(.))$ must satisfy assumption 1 and be such that the second order conditions of the principal's and the agent's programs hold. Some choices can be discarded according to these criterions and bounds on parameters of interest may be obtained through these constraints (see Salanié, 2005b, for an approach of this kind).

On the other hand, if one of the three structural functions is known, we can recover the other functions of interest. In particular, and contrarily to the previous general case, the knowledge of $\widetilde{S}(.)$ enables to recover the cost function everywhere. Another application of our result is regulation with ex-post observable costs. Suppose indeed that $\theta(y)C(y)$ is observed. Because $\theta(y)C'(y) = t'(y)$ is also identified, $C'/C(.)$ is identified. Then C can be recovered up to a multiplicative constant, which is given by the normalization $\theta(y_0) = \theta_0$. As a consequence, F_θ and $\widetilde{S}$ are identified. This result is in line with Perrigne and Vuong (2004)'s one.[6]

We review in the following several classical settings where this model is useful and show how identification can be obtained.

---

[6]Actually, the problem of Perrigne and Vuong (2004) is more involved because they consider the regulation model of Laffont and Tirole (1986) where the regulated firm can make a costly and unobserved effort to reduce its cost. Thus, they have to deal both with adverse selection and moral hazard.

3.1.1 Quality and Price Discrimination

In Mussa and Rosen (1978), the principal is a firm that produces a good of quality q at a cost $H(q)$. Agents have heterogenous preferences for quality θ (distributed following F_θ) and have a utility $U = \theta q - t$ if they pay t for a good of quality q.

In this setting, the objective function of the principal is unknown and depends on $H(.)$ ($S(q, \theta) = H(q)$ in the previous notation) whereas the utility of the agents is specified ($C(q) = q$). Our proposition implies that this model is identified nonparametrically.[7]

The same model can be used to study nonlinear pricing by a monopoly (Maskin and Riley, 1984). Our result shows that if the utility function of agents is specified, the model is identified.

3.1.2 Financial contracts

In Freixas and Laffont (1990) framework, the principal is a lender who provides a loan y to a borrower and has a utility $S(y) = t - Ry$ where R is the risk-free interest rate. Agents are firms with profit $U = \theta f(y) - t$. $\theta f(y)$ is the production of the firm, y represents the units of capital and θ is a productivity index.

Here, the objective function of the principal is known because R, the risk-free interest rate is observed. Our proposition implies that the production function $f(.)$ and the distribution of the types $F_\theta(.)$ are identified.

3.1.3 Regulation

In the Baron and Myerson (1982) model, the regulator maximizes a weighted sum of the consumers' surplus and the regulated firms defined by heterogenous cost functions of the form $\theta C(y)$. In our notation, we have:

$$S(y, \theta) - t(y) = (1 - \alpha) \left[\int_0^y p(u) du - t(y) \right] + \alpha \left[t(y) - \theta C(y) \right] \qquad (3.4)$$

From this equation, we derive $\widetilde{S}(y) = (1 - \alpha) p(y) + \alpha t'(y)$. Hence, when the price function $p(y)$ is observed (which is usually the case in the empirical literature), $\widetilde{S}(.)$ is known up to the parameter α and our proposition proves that $C(.)$ and $F_\theta(.)$ are identified up to this parameter.

---

[7]In this example, the principal minimizes his objective function instead of maximizing it.

3.2 Identification under exclusion restrictions

3.2.1 The setting

Proposition 3.1 states that the model is identified provided that one of the three structural functions is known. This condition may however be restrictive in certain settings. In the regulation problem, one has to assume that the weighting parameter α is known. In the price discrimination case, identification is based on the assumption of linearity of U in q. Moreover, one of the major empirical question in contract theory is whether the observed contracts are optimal compared to the theoretical ones (Chiappori and Salanié, 2002). Answering this question implies estimating $\widetilde{S}$ and compare it with the theoretical one. In this case, one can obviously not rely on the theoretical $\widetilde{S}$ to identify the model.

In this subsection, we examine the identification of the model when several menus of contracts are available but none of the functions $(F_\theta, C', \widetilde{S})$ is known. More precisely, we suppose that we observe variations in the menus of contracts under the exclusion restriction that the utility function and the distribution function of θ are not affected by these changes. We suppose in particular that there is no selection effect (see the discussion on this issue in subsection 3.4).

Exogenous variations in the menus can be observed for different reasons. The first and ideal case is experiments: if different contracts are proposed to people in a random way, then endogenous changes or selection problems are not a concern. For instance, the Rand Health Insurance experiment (see Manning et al., 1987) randomly assigned families who participate in the experiment to 14 different insurance plans. Similarly, Ausubel (1999) analyses the market for bank credit by using randomized mailed solicitations. The propositions vary in the interest rates and the duration of the loan. The econometrician may also use natural experiments where the objective function S of the principal changes for an exogenous reason.[8] Laws modifications are often good candidates for this purpose. Many examples have been already studied in the literature, especially in moral hazard situations (see e.g. Dionne and Vanasse, 1996; Chiappori et al., 1998; Chiappori et al., 1998; Banerjee et al., 2002). In the regulation context, one could also use changes in the government, which may induce variation of the parameter α, while the structural parameters (F_θ, C') remain constant (Gagnepain and Ivaldi, 2007). In a monopoly price discrimination model, the price of one input may increase, inducing a change in the cost function of the monopoly

---

[8]Our analysis could be adapted to the case where C' (resp. F_θ) alone varies, $(\widetilde{S}, F_\theta)$ (resp. $(\widetilde{S}, C')$) remaining constant. We focus on the variations of the principal's objective function because we believe it to be the most common situation.

and thus in S. However, in a partial equilibrium framework, this increase does not affect the utility function of the customer. In the delegation of production to agents, the firm may restructure their wage schedule for an exogenous reason such as, for instance, a managerial change. Another example on the French National Institute will be developed in section four (see also Lazear, 2000).

In the sequel, we denote by λ the index of these different menus. We suppose that there are K different indices and we denote their set by $\Lambda = \{\lambda_1, ..., \lambda_K\}$. $t(., \lambda)$ (resp. $\widetilde{S}(., \lambda)$) is the transfer function (resp. the marginal objective function at the optimum) corresponding to λ. The production chosen at equilibrium also depends on λ and we denote it by $y(\theta, \lambda)$, its distribution function being $F_y(., \lambda)$. Its inverse function is $\theta(y, \lambda)$. We now suppose that the econometrician has access to data of the form $(t(y_i, \lambda_i), y_i, \lambda_i), i \in \{1, ..., N\}$. Here, λ_i does not necessarily have an economic meaning: observing λ_i only indicates that the type of contract of individual i is known. As previously, $t(., \lambda)$ and $F_y(., \lambda)$ are identified for all $\lambda \in \Lambda$, and our aim is to recover the structural parameters $\left(C', F_\theta, (\widetilde{S}(., \lambda))_{\lambda \in \Lambda}\right)$.

We start from the first order condition and the monotonicity condition:

$$\frac{\partial t}{\partial y}(y, \lambda) = \theta(y, \lambda)C'(y) \tag{3.5}$$

$$F_y(y(\theta, \lambda), \lambda) = 1 - F_\theta(\theta) \tag{3.6}$$

These two equations, together with (2.3), show that it suffices to identify $(y(., \lambda))_{\lambda \in \Lambda}$ or equivalently $(\theta(., \lambda))_{\lambda \in \Lambda}$ to recover the structural parameters. The idea is thus to focus on these functions in order to derive our identification results. To do so, we first need to introduce two types of transforms that will be at the basis of our identification method.

3.2.2 The horizontal and vertical transforms

Equation (3.6) implies that for all $\theta \in \Theta$,

$$F_y(y(\theta, \lambda_i), \lambda_i) = F_y(y(\theta, \lambda_j), \lambda_j).$$

The ranks of $y(\theta, \lambda_i)$ and $y(\theta, \lambda_j)$ are identical in their respective distribution. Hence, letting $H_{ij}(y) = F_y^{-1}[F_y(y, \lambda_i), \lambda_j]$ denote the quantile-quantile transformation between the distribution of $y(\theta, \lambda_j)$ and $y(\theta, \lambda_i)$, we get

$$y(\theta, \lambda_j) = H_{ij}(y(\theta, \lambda_i)) \tag{3.7}$$

The horizontal transforms on figure 1 are identified and we can recover point (1) for instance if we know point (0).

Besides, let $\mathcal{Y}_i$ be the support of $y(\theta, \lambda_i)$. For $i \neq j$, suppose that $\mathcal{Y}_i \cap \mathcal{Y}_j \neq \emptyset$ and let $y \in \mathcal{Y}_i \cap \mathcal{Y}_j$. The first order condition implies that

$$\frac{\frac{\partial t}{\partial y}(y, \lambda_i)}{\theta(y, \lambda_i)} = \frac{\frac{\partial t}{\partial y}(y, \lambda_j)}{\theta(y, \lambda_j)}.$$

If we define the vertical transform $V_{ij}(.,.)$ by $V_{ij}(\theta, y) = \partial t / \partial y(y, \lambda_j) \times \theta / \partial t / \partial y(y, \lambda_i)$, we get

$$\theta(y, \lambda_j) = V_{ij}(\theta(y, \lambda_i), y) \tag{3.8}$$

Because $V_{ij}(.,.)$ is identified on $\mathbb{R} \times \mathcal{Y}_i \cap \mathcal{Y}_j$, the knowledge of $\theta(y, \lambda_i)$ implies the knowledge of $\theta(y, \lambda_j)$. In particular, it suffices to identify $\theta(., \lambda_1)$ to recover the other functions $\theta(., \lambda_k)_{2 \leq k \leq K}$. Starting from point (1) for instance on figure 1, we can identify point (2).

To conclude, starting from $(y_0, \theta(y_0, \lambda_1))$, we can identify $(y_1, \theta(y_1, \lambda_1))$ where $y_1 = H_{12}(y_0)$ and $\theta(y_1, \lambda_1) = V_{21}(\theta(y_0, \lambda_1), y_1)$. By induction, we identify all the black points in figure 1.
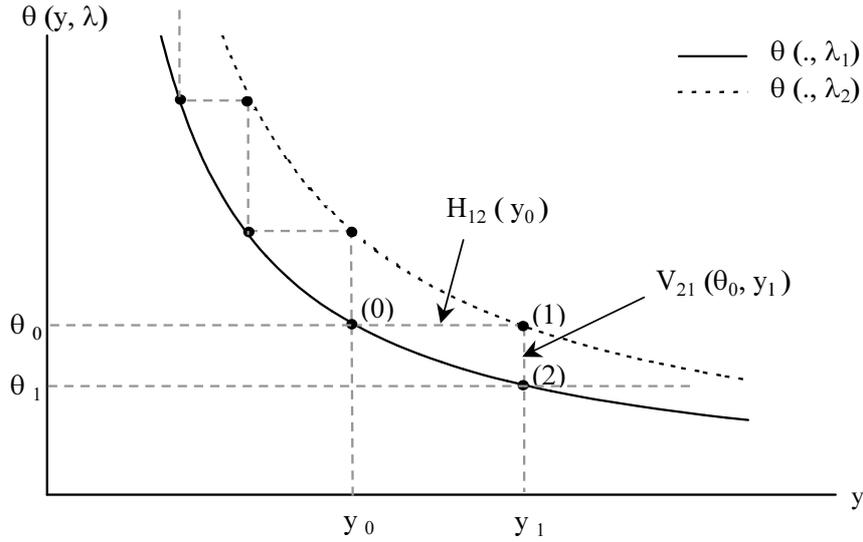

Figure 1: The horizontal and vertical transforms.

3.2.3   Identification results for $K = 2$.

Figure 1 corresponds to a situation where $\theta(y, \lambda_1) < \theta(y, \lambda_2)$ for all $y \in \mathcal{Y}_1 \cap \mathcal{Y}_2$. This does not hold in general, since $\theta(., \lambda_1)$ and $\theta(., \lambda_2)$ may cross. Actually, because $\theta(y, \lambda_1) < \theta(y, \lambda_2)$ is equivalent to $\partial t / \partial y(y, \lambda_1) < \partial t / \partial y(y, \lambda_2)$, the two cases can be distinguished by the data. As they lead to different results in terms of identification, we consider them separately.

Figure 1 suggests that without crossing, $\theta(., \lambda_1)$ can be identified on some points but not everywhere. This implies partial identification of the model. Theorem 3.2 formalizes this idea.

Theorem 3.2 *Suppose that $K = 2$, assumptions 1-3 hold and $\partial t / \partial y(., \lambda_1) < \partial t / \partial y(., \lambda_2)$. Then $C'(.)$ and $F_\theta(.)$ are identified on two sequences. Upper and lower bounds can be obtained elsewhere. The functions $\left(\widetilde{S}(., \lambda)_{\lambda \in \Lambda}\right)$ are not identified.*

Bounds on C' and F_θ can be recovered when $\partial t / \partial y(., \lambda_1) < \partial t / \partial y(., \lambda_2)$. Full nonparametric identification is not achieved but what is recovered can be sufficient to test parametric restrictions on F_θ or C'. In contrast with these positive results, the proposition shows that nothing can be learnt on the principal's value function when $K = 2$. Recovering $\widetilde{S}$ is indeed more demanding than identifying (C', F_θ) on some points, since it requires to recover f_θ (see equation (2.3)). Here, f_θ is unidentified because only isolated points of F_θ can be obtained.[9]

To recover full nonparametric identification, more structure on the model and on the functions of interest must be imposed. However, we need not fix or observe entirely one of the structural functions anymore as in the case $K = 1$. Usually, a parametric restriction on one of these functions will be sufficient to identify (and even overidentify) the model. Consider for instance the case of regulation, where the exogenous change of the transfer function is due to a modification of the unknown parameter α in (3.4) and let us call α_1 (resp. α_2) this parameter in the first (resp. second) sample. On the one hand, using proposition 3.2, $C'(.)$ is identified on a sequence $(y_n)_{n \in \mathbb{Z}}$. On the other hand, fixing $\alpha = \alpha_1'$ is equivalent to specifying the principal objective function. This defines, by proposition 3.1, a unique function $C'_{\alpha_1'}$. In general, the sequences $(C'(y_n))_{n \in \mathbb{Z}}$ and $(C'_{\alpha_1'}(y_n))_{n \in \mathbb{Z}}$ are equal for a unique value α_1', which ensures the identification of this parameter. We identify similarly α_2. By proposition 3.1, the whole model is identified.

We now turn to the case where functions $\partial t / \partial y(., \lambda_1)$ and $\partial t / \partial y(., \lambda_2)$ cross. In this case, the model can be fully recovered thanks to the intersection point. The proof is quite different from previously and can be explained as follows. By the normalization, an intersection point (y_c, θ_c) can always be fixed. For any y_0, define the sequence $(\theta_n)_{n \in \mathbb{N}}$ as in figure 2 from a given θ_0. We show that $(\theta_n)_{n \in \mathbb{N}}$ converges to θ_c if and only if $\theta_0 = \theta(y_0, \lambda_1)$. This enables to recover θ_0, since θ_c is known. Hence, $\theta(., \lambda_1)$ is fully identified and all the structural functions can be recovered.

---

[9]To obtain partial identification, more structure is needed. For example, if $S(y, \theta) = S(y)$, we can show that bounds can be recovered for $S(y_{n+1}) - S(y_n)$, where $(y_n)_{n \in \mathbb{N}}$ is the sequence of points defined as in figure 1.
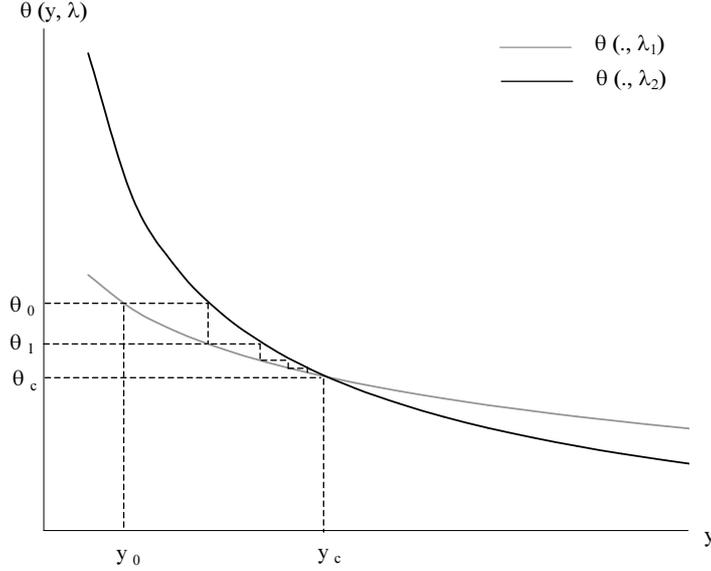
Figure 2: Identification when $\partial t/\partial y(., \lambda_1)$ and $\partial t/\partial y(., \lambda_2)$ cross.

Theorem 3.3 *Suppose that $K = 2$, assumptions 1-3 hold and $\partial t/\partial y(y_c, \lambda_1) = \partial t/\partial y(y_c, \lambda_2)$ for a given y_c in the interior $\overset{\circ}{\mathcal{Y}}_1$ of $\mathcal{Y}_1$. Then C', F_θ, $\widetilde{S}(., \lambda_1)$ and $\widetilde{S}(., \lambda_2)$) are identified on their support.*

Theorem 3.3 is reminiscent of the result of Guerre, Perrigne and Vuong (2005) in the context of first-price auctions with risk averse bidders. They also use exogenous variations (namely, the variation in the number of bidders) to obtain identification of the model at the limit, using a converging sequence (see their proposition 1).

3.2.4 Identification results for $K \geq 3$.

In this subsection, we study the noncrossing case with two or more exogenous changes. Here we have in hand not only the transforms H_{12} and V_{12}, but also (when $K = 3$) H_{13}, H_{23}, V_{13} and V_{23}. As a consequence, the set on which $\theta(., \lambda_1)$ is identified is larger. Figure 3 gives an example where starting from (y_0, θ_0) on the curve $\theta(., \lambda_1)$, we can identify $\theta(., \lambda_1)$ on y_1 as previously but also between y_0 and y_1 (on y_2 for instance). Proposition 3.4 defines the precise set where the function $\theta(., \lambda_1)$ is point identified.

Proposition 3.4 *Suppose that $K \geq 3$, assumptions 1-3 hold and $\partial t/\partial y(., \lambda_1) < ... < \partial t/\partial y(., \lambda_K)$. $\theta(., \lambda_1)$ is identified on $\overline{\mathbb{Y}} \cap \mathcal{Y}_1$ where $\overline{\mathbb{Y}}$ is the closure of the set $\mathbb{Y}$ defined by:*

$$\begin{cases} y_0 \in \mathbb{Y} \\ \text{For all } (y, i, j) \in \mathbb{Y} \times \{1, ..., K\}^2, \ y \in \mathcal{Y}_i \text{ implies } H_{ij}(y) \in \mathbb{Y} \end{cases}$$
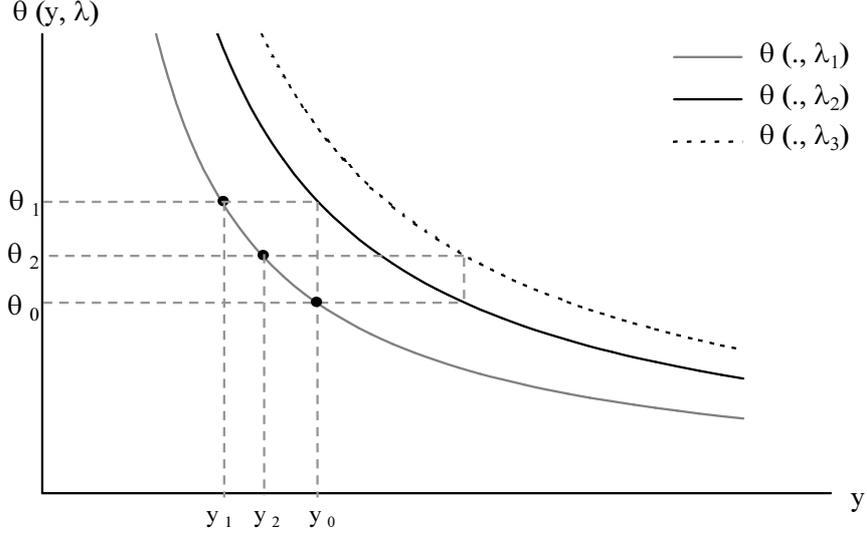
14

Figure 3: Identification with $K = 3$ different transfer functions.

It seems difficult to characterize $\mathbb{Y}$ more precisely without further restrictions. It may happen that $\overline{\mathbb{Y}} \cap \mathcal{Y}_1 \neq \mathcal{Y}_1$, so that similarly to the case $K = 2$, only bounds can be obtained on C' and F_θ. However, under the assumptions below, we prove that $\mathbb{Y}$ is dense in $\cup_{i=1,..,K} \mathcal{Y}_i$ and that the model is fully identified.

Assumption 4 *(separability in the transfer function) For all (y, λ),*

$$\partial t / \partial y(y, \lambda) = l(\lambda) m(y).$$

Assumption 5 *(non periodicity): there exists $1 \leq i < j < k \leq K$ such as*

$$\frac{\ln(l(\lambda_i)/l(\lambda_j))}{\ln(l(\lambda_k)/l(\lambda_j))} \notin \mathbb{Q}. \tag{3.9}$$

Up to ignoring some elements of Λ, we can let without loss of generality $i = 1$, $j = 2$ and $k = 3$.

Assumption 6 *(large support): $H_{21}(\mathcal{Y}_1 \cap \mathcal{Y}_3) \cap \mathcal{Y}_2 \neq \emptyset$ and $H_{23}(\mathcal{Y}_1 \cap \mathcal{Y}_3) \cap \mathcal{Y}_2 \neq \emptyset$.*

Assumption 5 is a technical condition which ensures that the identifying sequences are not periodic. Because almost every real are irrational, this assumption should not be seen as restrictive. The large support condition ensures that enough horizontal transforms can be performed to obtain new points where $\theta(., \lambda_1)$ is identified. This assumption can be easily checked in the data. The more restrictive assumption is the separability hypothesis. It

includes nevertheless the case of constant marginal transfers and is directly testable in the data.[10] Theorem 3.5 shows that the model is fully nonparametrically identified under these assumptions.

Theorem 3.5 *If $K \geq 3$ and assumptions 1-6 hold, $C'(.)$, $F_\theta(.)$ and $\widetilde{S}(., \lambda_i)$, for all $i = 1, .., K$, are identified on their support.*

Theorem 3.5 implies that when $K \geq 4$, the model is actually overidentified.[11] Indeed, we can recover (under suitable adaptations of the assumptions) F_θ and C' by using different subsets of Λ. If the different corresponding functions do not coincide, the model is rejected.

3.2.5 Identification results with continuous variations.

It may happen that there is actually a continuum of exogenous variations. In the case of price discrimination for instance, the prices of the input of the monopoly may take any value in an interval, implying also that the value function of the principal changes continuously. As mentioned above, the model with cost separability is now overidentified. Actually, identification can be obtained without this assumption.

Proposition 3.6 *Suppose that $\Lambda = [\underline{\lambda}, \overline{\lambda}]$, assumption 1 and 2 hold, θ is uniformly distributed on $[0, 1]$ (without loss of generality) and $\partial^2 t / \partial y \partial \lambda(y, \lambda) > 0$. Then*

- *$\partial C / \partial y(., .)$ is identified on $\{(y, \theta)/\theta \in \Theta, \ \exists \lambda \in \Lambda/\theta(y, \lambda) = \theta\}$*

- *$\widetilde{S}(., \lambda)$ is identified on $\{y/\exists(\theta, \lambda) \in \Theta \times \Lambda/\theta(y, \lambda) = \theta\}$.*

3.3 Testing the model

The empirical foundations of theoretical models from contract theory have been much debated in the literature. In this subsection, we identify three tests of our model, which correspond to well defined economic questions.

3.3.1 Do people react to incentives ?

Incentives are at the core of economic reasoning and many papers have sought evidence on the reactions to these incentives (see Prendergast, 1999, and Chiappori and Salanié, 2002,

---

[10]Note that it automatically implies $\partial t / \partial y(., \lambda_1) < ... < \partial t / \partial y(., \lambda_K)$, up to a reindexation.

[11]Whether the model is just identified or overidentified when $K = 3$ is unclear to us.

for surveys on this issue).[12] In our model and with $K \geq 2$, we can test such reactions by checking if

$$\partial t / \partial y(y, \lambda_1) < \partial t / \partial y(y, \lambda_2) \Longleftrightarrow \theta(y, \lambda_1) < \theta(y, \lambda_2)$$

holds for all $y \in \mathcal{Y}_1 \cap \mathcal{Y}_2$. Using (3.6), this condition may be written as

$$\left\{ y / \frac{\partial t}{\partial y}(y, \lambda_1) < \frac{\partial t}{\partial y}(y, \lambda_2) \right\} = \left\{ y / F_y(y, \lambda_1) > F_y(y, \lambda_2) \right\}. \tag{3.10}$$

In particular, when $\partial t / \partial y(., \lambda_1) < \partial t / \partial y(., \lambda_2)$, the model implies that $F_y(., \lambda_1)$ is stochastically dominated at the first order by $F_y(., \lambda_2)$. This implication can be straightforwardly tested by the data, for instance through a one-sided Kolmogorov-Smirnov test.

3.3.2 Are contracts optimal ?

Another important empirical issue of contract theory is the optimality of observed contracts. Nonstructural approaches can only bring limited clues on this issue and the structural papers have relied so far on parametric forms (see e.g. Ferrall and Shearer, 1999). In such a framework, the rejection of the model can either discard the parametric hypotheses or the theoretical framework. On the contrary, our nonparametric approach enables to test the theory solely.

The previous results imply that the model is fully or partially identified when $K \geq 2$ and without the knowledge of the theoretical value function of the principal. Hence, with such an extra assumption, the model is generally overidentified, and we can test whether this theoretical form is coherent with the data.

3.3.3 Does asymmetric information really matter ?

Lastly, a large literature has dealt with the empirical relevance of asymmetries of information (see Chiappori and Salanié, 2002, for a review). Generally speaking, the results of this literature throw doubts on the importance of such asymmetries. In this subsection, we question the possibility of testing nonparametrically the complete information model described in section 2.1 versus the asymmetric information model, when the information set of the principal is unknown. Proposition 3.7 sums up our findings.

Proposition 3.7 *Suppose that assumption 1, 2 and 3 hold. In general, no test of complete versus asymmetric information can be done if $K = 1$, even if one of the structural functions is known. When $K \geq 2$, a test can be implemented under one of the following conditions:*

---

[12]Overall, the conclusion of this research is rather positive.

- $$\left\{ y / \frac{\partial t}{\partial y}(y, \lambda_1) < \frac{\partial t}{\partial y}(y, \lambda_2) \right\} \neq \left\{ y / t(y, \lambda_1) < t(y, \lambda_2) \right\} \tag{3.11}$$

- *assumption 4 holds and C' is known ;*

- *assumption 4 holds, one function $\widetilde{S}(., \lambda_0)$ is known and $y \mapsto t(y, \lambda_0) \times (1 - F_y(y, \lambda_0))$ is not constant.*

Hence, the possibilities of testing for asymmetric information are rather limited. In particular, the two models cannot be distinguished under assumption 4 without auxiliary information, even when $K \geq 2$. This stems from the fact that under this assumption, the two models lead to the same function $\theta(., \lambda_1)$.[13] This result contrasts with the previous literature on insurance in which nonstructural tests are performed using auxiliary variables (Puelz and Snow, 1994; Chiappori and Salanié, 2000).

3.4 Discussion and extensions

In this subsection, we come back on assumptions 2 and 3. We also discuss the issue of selection effects.

3.4.1 Bunching

We have maintained up to now the assumption that no bunching occurs at the equilibrium. As mentioned previously, this assumption is testable, by checking if F_y admits a mass point. Bunching is also equivalent to $t(.)$ being continuously differentiable. Many observed contracts do not fulfill this requirement. For instance many production contracts exhibit kinks (see e.g. Ferrall and Shearer, 1999). In this case, bunching provides more information on the model. Figure 4 gives the intuition for this result. Fixing (θ_0, y_0) at one extremity of the bunching, it is possible to identify $\theta_1 = (\partial t / \partial y^-(y_0, \lambda_1))/(\partial t / \partial y^+(y_0, \lambda_1))\theta_0$ i.e. the interval $[\theta_0, \theta_1]$. The horizontal and vertical transformations of this interval allows the econometrician to identify $\theta(., \lambda_1)$ and consequently $C'(.)$, $F_\theta(.)$ and the functions $(\widetilde{S}(., \lambda))_{\lambda \in \Lambda}$ on several segments, even when $K = 2$.

3.4.2 The false moral hazard model

Our results are based on the cost separability assumption $C(\theta, y) = \theta C(y)$. Other restrictions are however possible. One example is given by the "false moral hazard" model (see

---

[13]The overidentification restrictions on $(\theta(., \lambda))_{\lambda \in \Lambda}$ (when $K \geq 4$) are useless to test between the two models because if they fail to hold, both models are rejected.
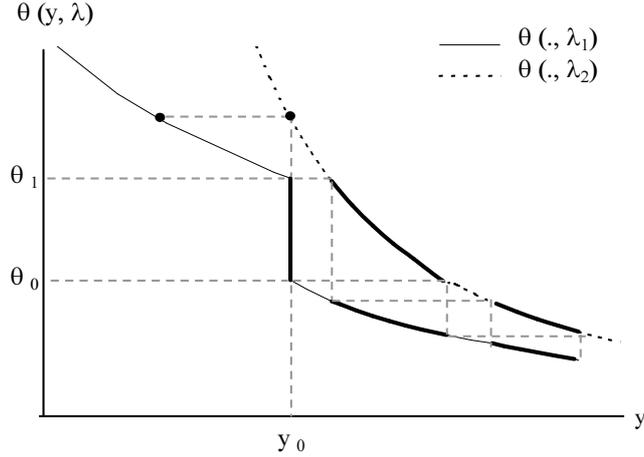
Figure 4: Identification with bunching. Thick lines (resp. black points) correspond to the identified intervals (resp. points) of $\theta(., \lambda_1)$ and $\theta(., \lambda_2)$.

Laffont and Martimort, 2002, p. 287). In this framework, we suppose that the production of the agent depends on θ but also on the level of his effort e, so that $y = g(\theta, e)$. The cost $C(e)$ depends only on the effort e. The agent observes the random term θ before he chooses his effort so that he maximizes $t(g(\theta, e), \lambda) - C(e)$. Though apparently close to the moral hazard model, this model does not share its properties (the trade-off between efficiency and risk insurance especially) and is rather an adverse selection model.

From an econometric point of view, the identification of the false moral hazard model is very close to the previous analysis. The model satisfies $C(y, \theta) = C(g_2^{-1}(\theta, y))$ where $g_2^{-1}(\theta, .)$ is the inverse function of $g(\theta, .)$. As previously, the structural parameters are C', F_θ and the $(\widetilde{S}(., \lambda))_{\lambda \in \Lambda}$. In this framework, proposition 3.1 and 3.4 are still valid. The equivalent of proposition 3.5 also holds if the marginal transfer is constant and the production function is separable, $g(\theta, e) = h_1(\theta)h_2(e)$, as with Cobb-Douglas functions.

3.4.3 Selection effects

We have supposed until now that variations in the transfer functions do not yield any changes in F_θ. However, selection effects can be important. Lazear (2000), for instance, showed that half of the productivity increase observed in a car glass company after moving from constant wages to piece rates could be explained by the arrival of more productive worker. These effects are not taken into account in our model where all types of agent participate. Hence, our analysis is not valid in general when selection occurs.

There is, however, a simple way to detect selection effects when panel data are available. It suffices indeed, as in Lazear (2000) to compare the distributions of the production of

the stayers and the entrants. If these distributions differ significantly, selection should be dealt with by modeling the selection process.

4 Application

In this section, we apply our results on contract data between the French National Institute of Economics and Statistics (Insee) and the interviewers it hires to make household surveys. For each survey, Insee's interviewers receive a random sample of households close to their residence and has to fulfill a maximum number of face to face interviews. The Insee cannot compel its interviewers to obtain a given number of interviews and it provides them incentives through a linear scheme. Each interviewer receives a basic wage plus a bonus for each interview he achieves.

This application contributes to the literature of provisions of incentive by firms (see Prendergast, 1999, for a survey) which has been interested in studying 1) to what extent agents react to incentives and 2) the optimality of the observed contracts. Our approach is structural and similar to Paarsch and Shearer (2000) and Ferrall and Shearer (1999), but nonparametric. In these models, the production of a worker is a known function of his effort and his type. As explained in subsection 3.4, they are what Laffont and Martimort (2002) call "false moral hazard" models. These models have a structure that are very similar to adverse selection models and our results apply.

Here, for the ease of the presentation, we model the behavior of interviewers without making explicit reference to the underlying effort they produce. In a survey $j \in \{1, ..., J\}$, an interviewer $i \in \{1, ..., N\}$ receives a random sample of housings. The households to interview may be easy or difficult to contact and reluctant or not to accept the questionnaire. We summarize the average difficulty of this sample by a parameter $\theta_{ij} \in \mathbb{R}^+$ which is observed by the interviewer.[14] Letting y denote the response rate of the sample, we assume that the cost of interviewing the households (via his effort) is separable and writes as $\theta_{ij} C_j(y)$. The interviewer receives $w_j + \delta_j$ from Insee when the interview is achieved and w_j otherwise. The program of interviewer i is thus similar to the agent's program we have been looking at and writes as

$$\max_y \delta_j y - \theta_{ij} C_j(y).$$

---

[14]Before trying to contact the households, interviewers must locate the housings of their sample (in order to identify unoccupied or destroyed housings, for instance). During this phase, the interviewer learns the difficulty θ_{ij} of his sample.

Finally, Insee is an institute depending on public money and maximizing the social value of each survey. We denote by λ_j the "price" of the information contained in a household's answers, i.e. the social value of an interview in survey j. Hence, the Insee's objective function, associated with an interviewer whose response rate is y, writes as

$$S(y, \lambda_j) = \lambda_j y$$

4.1 The data

We use data on three household living conditions surveys ("enquête Permanente sur les Conditions de Vie des Ménages", PCV hereafter) which took place in October 2001, 2002 and 2003 ($j = 1, 2, 3$). In 2001 and 2002, the focus of the survey was put respectively on the use of new technologies and participation in associations, while in 2003, the survey studied education practices in the family. As a consequence, almost all households were eligible to the survey in 2001 and 2002 whereas only families were eligible in 2003. Otherwise, the surveys were identical in their designs and rules for the fieldwork.

In all surveys, we restrict our attention to the housings where more than three persons lived at the time of the 1999 census. This information is indeed available to the interviewer before conducting the surveys and is a good proxy for the eligibility of the household in 2003. For these households, the bonus for achieving an interview increased from 20 and 20.2 euros in 2001 and 2002 to 23.4 euros in 2003. To avoid selection effects, we also focus on the interviewers who conduct the three surveys.

We interpret this change as a modification of the principal objective function. We believe that the 2003 survey on education was considered by Insee to be more important than the ones on new technologies and participation to associations. Indeed, there is much debate in France on the relationship between families, education and the emergence of inequalities (see for instance the report of the Haut Conseil de l'Education in 2007 on this topic). More formally, more publications from Insee and other institutions were based on this survey and the questionnaire was slightly longer in 2003. Given these elements, we believe that the social value of an interview was higher in 2003 ($\lambda_1 = \lambda_2 < \lambda_3$). However, because the surveys were drawn in the same way, conducted in the same period, were close in time and had identical rules for the fieldwork, we assume that for all interviewers i, the distributions of θ_{ij}, $j \in \{1, 2, 3\}$ are identical. We also suppose that $C_1' = C_2' = C_3'$, even if an interview in 2003 was a little longer. Indeed, the cost of achieving an interview is mainly due to contact tries rather than the length of the interview itself.[15] Under these hypotheses, the

---

[15]This claim is supported by the discussions that we had with interviewers.

variation in the transfer function is exogenous, as defined in subsection 3.2.

	Interviewers	Bonus	Response rates	
			Mean	Std error
2001	236	20	79.6%	0.188
2002	236	20.2	80.4%	0.199
2003	236	23.4	83.8%	0.144

Table 1: Descriptive statistics in 2001, 2002 and 2003 surveys.

Our data consists in the identification number of the interviewer and his response rate, for each survey. The response rate of an interviewer is defined as the ratio of the number of respondents on the number of housings which are in the field of the survey (i.e., excluding secondary, unoccupied and destroyed housings for instance). Table 1 summarizes the main information about the surveys. It appears that the 2001 and 2002 surveys are very similar and we aggregate them in the rest of the application to obtain more precise results.[16] The 2003 survey is also similar to the other ones except for a higher response rate. Figure 5 displays more precisely the distribution function of the response rates for the 2001-2002 and 2003 surveys. As predicted by the theory, the distribution function of the 2003 survey stochastically dominates the one of 2001-2002, which proves that interviewers react to incentives.[17] We find that on average production increases by 5% when the piece rate increases by 16%. This result suggests a significant incentive effect, but relatively smaller than what the previous literature has found (Lazear, 2000; Paarsch and Shearer, 2000).

Figure 5 also displays several jumps in the distribution functions. We do not interpret them as evidence of bunching as the transfer functions do not exhibit kinks. These mass points are rather due to finite approximations of the response rates, and we neglect the error term in our estimation.

---

[16]The difference between the two average response rates is not significant at 5%, contrarily to differences between 2001 (or 2002) and 2003. In the following, we suppose that the two bonuses were identical and equal to 20.2 euros.

[17]The one-sided Kolmogorov-Smirnov test which tests the equality of the distribution functions rejects the null hypothesis at 5%.
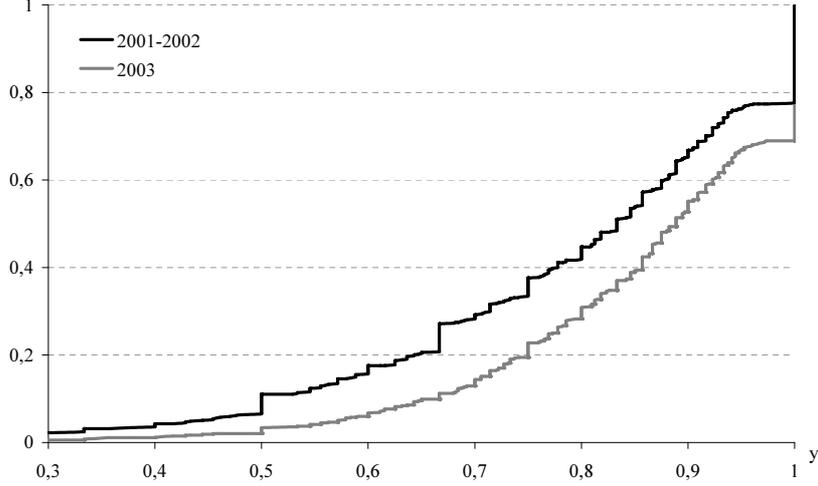
Figure 5: Distribution functions of the response rates for 2001-2002 and 2003 surveys.

4.2 Estimation

In this subsection, we define and study the estimators of the structural parameters in the framework of our application. We suppose that $K = 2$ and assume that $t(y, \lambda_j) = \delta_j y$.[18] Our asymptotic results rely on a standard assumption on the sample $(y_{ij})_{i \in \{1,...,N\}, j \in \{1,2\}}$, which does not impose any dependency structure between θ_{i1} and θ_{i2}.[19]

Assumption 7 *(independent sampling)* $(\theta_{11}, ..., \theta_{N1})$ *(resp.* $(\theta_{12}, ..., \theta_{N2})$*) are independently drawn according to* F_θ.

By proposition 3.2, we identify the function $\theta(., \lambda_1)$ on an increasing sequence $(y_n)_{n \in \mathbb{Z}}$. More precisely (see equation (6.5) in the proof of proposition 3.2), we identify the sequence

$$y_n = H_{12}^n(y_0) \mathbb{1}_{H_{12}(y_{n-1}) \in \mathcal{Y}_1} + y_{n-1} \mathbb{1}_{H_{12}(y_{n-1}) \notin \mathcal{Y}_1}, \quad n \in \mathbb{Z},$$

where $H_{12}^n = H_{12} \circ ... \circ H_{12}$ for any $n \in \mathbb{N}$ (and similarly, $H_{12}^{-n} = H_{12}^{-1} \circ ... \circ H_{12}^{-1}$).

Moreover, using $\theta(H_{12}(y), \lambda_1) = V_{21}(\theta(y, \lambda_1), H_{12}(y))$ and by induction, $\theta_n = \theta(y_n, \lambda_1)$ satisfies:

$$\theta_n = (\delta_1/\delta_2)^n \theta_0 \mathbb{1}_{H_{12}(y_{n-1}) \in \mathcal{Y}_1} + \theta_{n-1} \mathbb{1}_{H_{12}(y_{n-1}) \notin \mathcal{Y}_1},$$

where θ_0 is chosen arbitrarily.

---

[18] Actually, our results hold for any transfer function.

[19] We suppose for simplicity that the same agents are observed in both contracts.

These sequences are estimated as follows. For $j \in \{1, 2\}$, let $\widehat{F}_j$ (resp. $\widehat{F}_j^{-1}$) denote the empirical distribution function (resp. empirical quantile function) of the $(y_{ij})_{i \in \{1,\ldots,N\}}$. Our estimator of H_{12} is

$$\widehat{H}_{12}(x) = \widehat{F}_2^{-1} \circ \widehat{F}_1(x).$$

For all $n \in \mathbb{Z}$, we estimate y_n by

$$\widehat{y}_n = \widehat{H}_{12}^n(y_0) \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}}_1} + \widehat{y}_{n-1} \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}}_1},$$

where $\widehat{\mathcal{Y}}_1 = [\min_i y_{i1}, \max_i y_{i1}]$, and θ_n by

$$\widehat{\theta}_n = (\delta_1/\delta_2)^n \theta_0 \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}}_1} + \widehat{\theta}_{n-1} \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}}_1}.$$

Now, let us now turn to F_θ and C'. These functions are point identified respectively on $(\theta_n)_{n \in \mathbb{Z}}$ and $(y_n)_{n \in \mathbb{Z}}$ and we focus on the estimation of $F_\theta(\theta_n)$ and $C'(y_n)$. First, using (3.6), $F_\theta(\theta_n)$ satisfies $F_\theta(\theta_n) = 1 - F(y_n, \lambda_1)$. Hence, we define $\widehat{F_\theta(\theta_n)}$ by

$$\widehat{F_\theta(\theta_n)} = 1 - \widehat{F}_1(\widehat{y}_n).$$

Similarly, by 3.5, $C'(y_n) = \delta_1/\theta_n$. Thus, we consider the following estimator:

$$\widehat{C'(y_n)} = \frac{\delta_1}{\widehat{\theta}_n}.$$

The following theorem establishes the consistency of $(\widehat{\theta}_n, \widehat{F_\theta(\theta_n)})$ and $(\widehat{y}_n, \widehat{C'(y_n)})$. The result relies on assumption 8, which prevents any y_n from being on the boundary of $\mathcal{Y}_1$. Since the measure of this boundary is zero, this assumption is unrestrictive.

Assumption 8 *For all $n \in \mathbb{Z}$, $y_n \notin \mathcal{Y}_1 \setminus \overset{\circ}{\mathcal{Y}}_1$.*

Theorem 4.1 *Suppose that $K = 2$, $t(y, \lambda_j) = \delta_j y$ and assumptions 1-3, 7-8 hold. Then, for all $n \in \mathbb{Z}$ and when $N \to +\infty$,*

$$(\widehat{\theta}_n, \widehat{F_\theta(\theta_n)}) \overset{\mathbb{P}}{\longrightarrow} (\theta_n, F_\theta(\theta_n))$$

$$(\widehat{y}_n, \widehat{C'(y_n)}) \overset{\mathbb{P}}{\longrightarrow} (y_n, C'(y_n))$$

By the monotonicity of F_θ and C', the estimators of $F_\theta(\theta_n)$ and $C'(y_n)$ enable us to build bounds on the two functions.

4.3 Results

We now apply this estimation method on our contract data. Firstly, starting from a middle point $y_0 = 0.6$ (with $\theta_0 = 1$), we estimate $(y_n, \theta_n, F_\theta(\theta_n), C'(y_n))_{n \in \mathbb{Z}}$ as indicated and obtained 12 distinct points which correspond to $n \in \{-3, ..., 8\}$. Figures 6 and 7 display the estimations of the bounds on $F_\theta(.)$ and $C'(.)$ respectively, and their 95% confidence interval obtained by bootstrap. With twelve points, the bounds on both functions are close and we are able to correctly retrieve their shape. The highly convex form of the cost function shows in particular that incentives are relatively large for small values of the production but sensitively lower for higher ones. This may explain the small average effect of incentives that we have found compared to the previous results of the literature. We also note that, as expected, the errors accumulate in the estimation procedure and that the width of the confidence intervals on the bounds of F_θ (resp. C') increases with $|\theta - 1|$ (resp. $|y - 0.6|$).
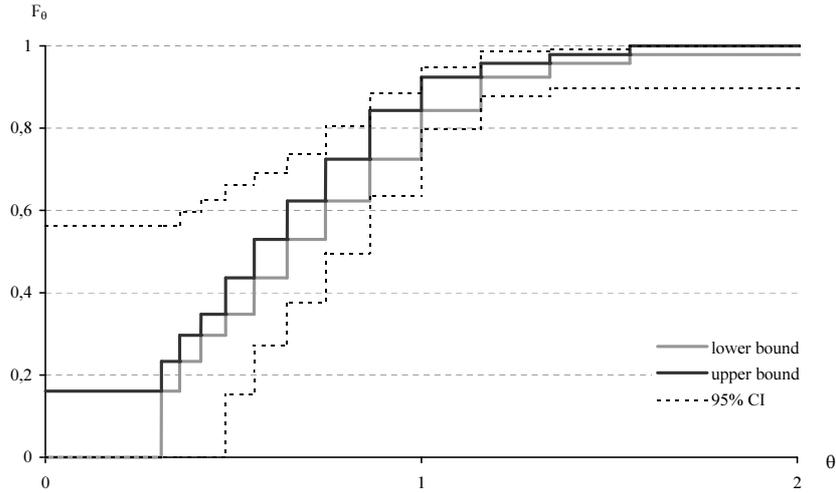

Figure 6: Estimated bounds on $F_\theta(.)$

Secondly, this information is used to estimate the objective function of the principal $S(y, \lambda_3) = \lambda_3 y$ in 2003. Using equation 6.4 in appendix A, it can be shown that λ_3 satisfies for all $n \neq 0$:

$$\lambda_3 = \delta_3 \left(1 + \frac{\ln(\theta_n/\theta_0)}{\ln(F_\theta(\theta_n)/F_\theta(\theta_0))} \right)$$

We then obtain eleven different estimators of λ_3, for $n \in \{-3, 8\}$, $n \neq 0$:

$$\widehat{\lambda}_{3n} = \delta_3 \left(1 + \frac{\ln(\widehat{\theta}_n/\theta_0)}{\ln(\widehat{F_\theta(\theta_n)}/\widehat{F_\theta(\theta_0)})} \right)$$
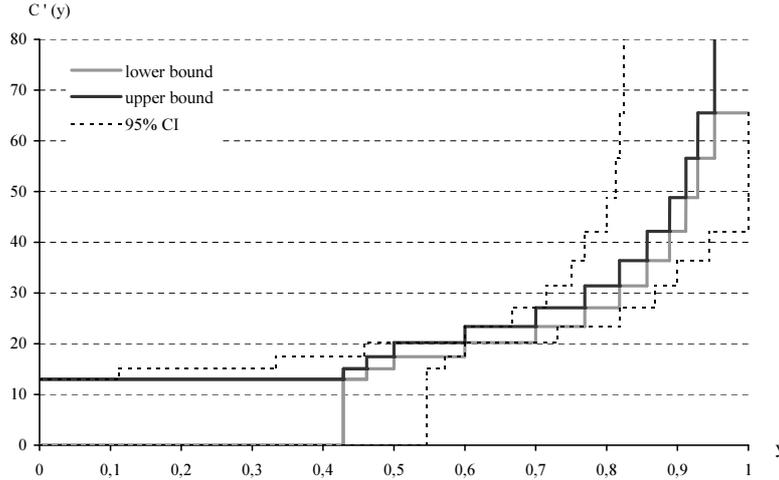
25

Figure 7: Estimated bounds on $C'(.)$.

and the model is overidentified (see the discussion following proposition 3.2). Figure 8 depicts these estimators and their associated confidence interval calculated by bootstrap. We see that $\widehat{\lambda}_{3n}$ does not appear to be constant in n, which contradicts the contracts' optimality. More formally, assuming asymptotic normality, we compute the Wald statistic of the test $\lambda_{-3} = ... = \lambda_8$. This statistic equals 65.5 and should be compared to the 95% quantile of a $\chi^2(11)$, i.e., 19.7. Hence, we clearly reject the contracts' optimality at the 5% level (p-value$\simeq 10^{-9}$).

The rejection of contracts' optimality is also reinforced by the violation of the Informativeness Principle which states that all factors correlated with performance should be included in the contracts (Prendergast, 1999). In our application, contracts are not optimal as the bonus does not depend for instance on the region in which interviewers are working, even if the average response rate in Paris area (0.73% in 2003) is significantly lower than in the rest of France (0.85%).

To understand why Insee designs linear contracts, it is interesting to measure the cost of using simple but inefficient incentives. We thus compare the actual surplus of the institute with the one it would obtain under the use of optimal contracts. We also compute the surplus of Insee under complete information and estimate the cost of asymmetric information. To do so, we need to recover λ_3 in the objective function of Insee. However, under the assumption that Insee maximizes his objective function only in the class of linear contracts, equation (3.3) is not valid anymore. It can be shown, in this case, that λ_3 is not identified even with $K = 2$. We thus adopt a parametric approach and impose a structure coherent with our previous nonparametric analysis. More precisely, we suppose
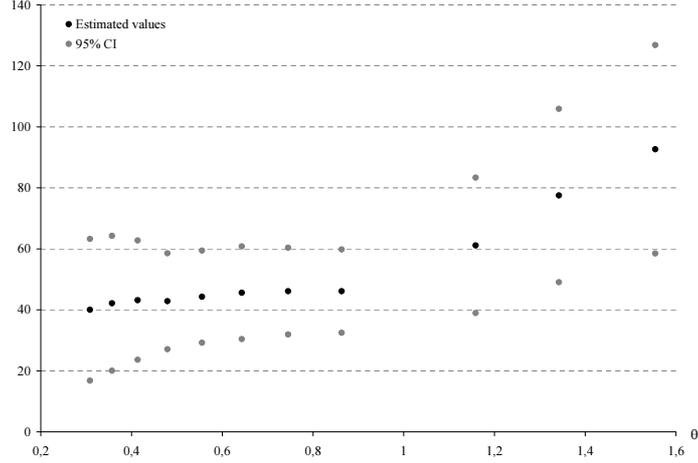
Figure 8: Estimated $\widehat{\lambda}_{3n}$, $n \in \{-3, 8\}$, $n \neq 0$.

that θ follows a Weibull distribution $F_\theta(\theta) = 1 - \exp(-a\theta^b)$ for all $\theta \in \mathbb{R}^+$ and that the cost function takes the form $C'(y) = \alpha (y/1 - y)^\beta$ on $[0, 1[$. The parameters of interest are estimated by regressing $\ln(-\ln(\widehat{F}_j(y)))$ on $\ln[(1 - y)/y]$ (see appendix B). We obtain $\widehat{a} = 1.87 \,(0.03)$, $\widehat{b} = 1.89 \,(0.04)$, $\widehat{\alpha} = 16.7 \,(0.07)$ and $\widehat{\beta} = 0.46 \,(0.01)$. Figure 9 shows that these parametric forms perfectly fits the nonparametric estimated points.
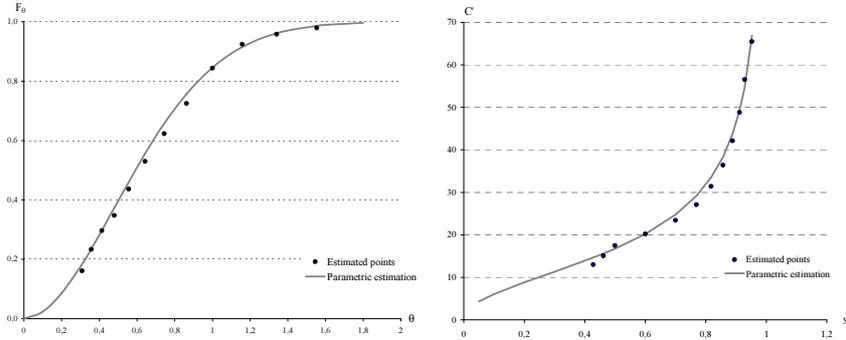

Figure 9: Parametric estimation of F_θ and C'.

Under these parametric restrictions, we are able (see appendix B) to estimate λ_3 and find that the social value of an interview to be $\widehat{\lambda}_3 = 102.1$ euros. We also compute the expected surplus under full information and the expected surplus under incomplete information with linear or optimal contracts. Table 2 summarizes our results. We find that the surplus loss associated with the use of linear contracts is about 9% and that the response rate also decreases by 9% compared to optimal contracts. This result contrasts with the previous literature in which linear contracts were thought to be quite inefficient. Ferrall and Shearer (1999), for instance, evaluate this loss at 50%. Our results point out on the contrary that

the cost is quite small and that optimal contracts are not highly nonlinear. This may explain why firms widely use linear contracts compared to nonlinear ones: they are less costly to implement and almost efficient.

Environment	Pay method	E[surplus]	Relative	E[response rate]
Full information	Optimal contract	86.0	1.00	0.99
Incomplete information	Optimal contract	72.5	0.84	0.92
Incomplete information	Linear contract	65.9	0.77	0.84

Table 2: Surplus and response rates under alternative compensation schemes.

Finally, we find moderate cost of incomplete information. The surplus under asymmetric information is 77% of what it could be under complete information. Two third of this loss is due to incomplete information whereas only one third is associated with the simple bonus system.

5   Conclusion

This work contributes to the recent structural analysis of incentive problems. First, by focusing on the general adverse selection model, we complement Perrigne and Vuong (2004)'s paper. Our result that these models are not fully identified is important to understand what restrictions are needed to recover the functions of interest in different settings such as regulation, nonlinear pricing or taxation. Second, we propose a way to exploit exogenous changes in order to identify or test nonparametrically the model and the contracts' optimality. The recursive method we develop enables us to analyze existing experiments or natural experiments. To our knowledge, such a recursive method is new. A consequence is that the econometric procedure, which is based on this method, is also a novelty. Third, studying the provision of incentives in firms, we test nonparametrically and reject the contracts' optimality proposed by Insee. We also estimate Insee's surplus to be 77% of the full information surplus. Two third of this loss correspond to the cost associated with asymmetric information, whereas the use of inefficient linear contracts only explains one third of it. This result may explain why firms widely use these contracts that are also very easy to implement.

Beyond these estimations, this approach can be useful to firms for determining the optimal contracts that they should implement. By proposing different contracts to random samples

of the population, the firm will learn the structural parameters of its agents and will then be able to design the optimal contract.

The paper also raises several challenging issues. Firstly, the properties of our estimators with three or more different contracts remain to be established. This may be difficult because identification is obtained at the limit with a density argument. Besides, one can wonder whether our strategy could be adapted to the moral hazard setting. Testable implications of this model have already been brought to light (Abbring et al., 2003, Chiappori et al., 2006), but its nonparametric identifiability has not been settled yet.

6   Appendix A: proofs

Proof of proposition 3.1

The first order condition of the agent writes

$$C'(y) = \frac{t'(y)}{\theta(y)} \tag{6.1}$$

Besides, by monotonicity of $\theta(.)$,

$$F_y(y) = 1 - F_\theta(\theta(y)) \tag{6.2}$$

By assumption 3, the first order condition of the principal writes as

$$\widetilde{S}(y) = \theta(y)C'(y) + \frac{F_\theta(\theta(y))}{f_\theta(\theta(y))}C'(y).$$

Thus, by (6.1),

$$\widetilde{S}(y) = \left[1 + \frac{F_\theta(\theta(y))}{\theta(y)f_\theta(\theta(y))}\right]t'(y).$$

Moreover, by (6.2), $F_\theta(\theta(y)) = 1 - F_y(y)$ and $f_\theta(\theta(y)) = -f_y(y)/\theta'(y)$. Thus,

$$\widetilde{S}(y) = \left[1 - \frac{1 - F_y(y)}{f_y(y)}\frac{\theta'}{\theta}(y)\right]t'(y). \tag{6.3}$$

Now, for any strictly decreasing and differentiating function $\theta(.)$, it is possible to define $C'(.)$, $F_\theta(.)$ and $\widetilde{S}$ using respectively equation (6.1), (6.2) and (6.3). For the model to be not identified, it is sufficient to prove that two sets of such functions satisfy assumption 1. Because assumption 1 is satisfied for the true function $\theta^0(.)$, we conclude that the assumption is satisfied also locally around $\theta^0(.)$. Hence, the model is not identified.

If $C'(.)$ (resp. $F_\theta(.)$) is known, equation 6.1 (resp 6.2) enables to identify $\theta(.)$ on $\mathcal{Y}$. Then, $F_\theta(.)$ (resp C') is identified on Θ (resp. $\mathcal{Y}$) using the previous equations. Finally, $\widetilde{S}$ is identified by (6.3).

Lastly, if $\widetilde{S}$ is known, (6.3) also writes as

$$\ln(\theta(y)) = \ln(\theta_0) + \int_{y_0}^{y} \left(1 - \frac{\widetilde{S}(u)}{t'(u)}\right) \frac{f_y(u)}{1 - F_y(u)} du \tag{6.4}$$

Hence $\theta(.)$ is identified. By (6.1) and (6.2), C' and F_θ are also identified. ∎

Proof of theorem 3.2

Firstly, we prove that $\theta(., \lambda_1)$ is identified on the closure of a sequence $(y_n)_{n\in\mathbb{Z}}$. If $\theta(y, \lambda_1)$ is known and if $H_{12}(y) \in \mathcal{Y}_1$ then $\theta(H_{12}(y), \lambda_1) = V_{21}(\theta(y, \lambda_1), H_{12}(y))$ is identified because H_{12} and V_{21} are.

Hence, using $y_0 \in \mathcal{Y}_1$, we deduce by induction that $\theta(., \lambda_1)$ is identified on the increasing sequence[20]

$$y_n = H_{12}^n(y_0)\mathbb{1}_{H_{12}(y_{n-1})\in\mathcal{Y}_1} + y_{n-1}\mathbb{1}_{H_{12}(y_{n-1})\notin\mathcal{Y}_1} \tag{6.5}$$

for any $n \geq 0$. Similarly, $\theta(., \lambda_1)$ is identified on the sequence

$$y_n = H_{21}^{-n}(y_0)\mathbb{1}_{H_{12}(y_{-n+1})\in\mathcal{Y}_2} + y_{-n+1}\mathbb{1}_{H_{12}(y_{-n+1})\notin\mathcal{Y}_2}$$

for any $n \leq 0$. By continuity of $\theta(., \lambda_1)$, the function is actually identified on the closure of $\{y_n, n \in \mathbb{Z}\}$.

Then, using the property that $\theta(., \lambda_1)$ is decreasing, we obtain that for any $y_{n-1} < x < y_n$, $\theta(y_{n+1}, \lambda_1) < \theta(x, \lambda_1) < \theta(y_n, \lambda_1)$ and similarly around the bounds.

Using this results and equations (6.1) and (6.2), we obtain that $C'(.)$ and $F_\theta(.)$ are identified respectively on the sequences $(y_n)_{n\in\mathbb{Z}}$ and $(\theta(y_n, \lambda_1)_{n\in\mathbb{Z}}$ whereas bounds are obtained for other values.

Now let us prove that $\theta(., \lambda_1)$ is not identified outside of the sequence $(y_n)_{n\in\mathbb{Z}}$. As in proposition 3.1, we derive from this result that $C'(.)$ and $F_\theta(.)$ are not identified outside the sequences $(y_n)_{n\in\mathbb{Z}}$ and $(\theta(y_n, \lambda_1)_{n\in\mathbb{Z}}$. Furthermore, $\theta'(., \lambda_1)$ is identified nowhere. By (6.3), this proves that the functions $\widetilde{S}(., \lambda_1)$ and $\widetilde{S}(., \lambda_2)$ are not nonparametrically identified.

---

[20] f^n denotes $f \circ ... \circ f$ and not $f \times ... \times f$ for any function f and $n \in \mathbb{N}$. Similarly, $f^{-n} = f^{-1} \circ ... \circ f^{-1}$.

To prove that $\theta(., \lambda_1)$ is not identified outside of the $(y_n)_{n \in \mathbb{Z}}$, we show that defining $\theta(., \lambda_1)$ on $\mathcal{Y}_1$ can be reduced by the first order conditions to defining this function on $]y_0, y_1[$ and that almost no restriction can be imposed on this interval.[21]

Suppose indeed that $\theta(., \lambda_1)$ is a known and differentiable function on the interval $]y_0, y_1[$. For any y in this interval, we are able to construct a sequence $(\tilde{y}_n)_{n \in \mathbb{Z}}$ with $\tilde{y}_0 = y$, just as we defined the sequence $(y_n)_{n \in \mathbb{Z}}$. Hence, $\theta(., \lambda_1)$ is defined on all these sequences i.e. on $\mathcal{Y}_1$. Furthermore, this function is differentiable everywhere except eventually at points $(y_n)_{n \in \mathcal{Z}}$. Using $\theta(H_{12}(y), \lambda_1) = V_{21}(\theta(y, \lambda_1), H_{12}(y))$, the differentiability of the function at these points is ensured as soon as

$$
\begin{aligned}
H'_{12}(y_0)\theta'_-(y_1, \lambda_1) &= H'_{12}(y_0)\theta'_+(y_1, \lambda_1) \\
&= H'_{12}(y_0)\frac{\partial V_{21}}{\partial y}(\theta(y_0, \lambda_1), y_1) + \theta'_+(y_0, \lambda_1)\frac{\partial V_{21}}{\partial \theta}(\theta(y_0, \lambda_1), y_1)
\end{aligned}
$$

where $\theta'_-(y_1, \lambda_1)$ (resp. $\theta'_+(y_1, \lambda_1)$) is the left (resp. right) derivative of $\theta(., \lambda_1)$ in $y = y_1$.

Finally, consider all the functions $\theta(., \lambda_1)$ defined on $]y_0, y_1[$ that can be extended to differentiable function on $\mathcal{Y}_1$. By construction, all these functions are coherent with respect to the horizontal and vertical transforms. Hence, $\theta(., \lambda_1)$ is not identified and proposition 3.2 follows. ∎

Proof of theorem 3.3

By the normalization and the fact that $y_c \in \overset{\circ}{\mathcal{Y}}_1$, we can always fix $0 < \theta_c < \infty$ such that $\theta_c = \theta(y_c, \lambda_1)$. Let y'_c denote the greatest intersection point of $\partial t/\partial y(., \lambda_1)$ and $\partial t/\partial y(., \lambda_2)$ which is smaller than y_c ($y'_c = \inf \mathcal{Y}_1$ if such a point does not exists). We suppose without loss of generality that $\partial t/\partial y(., \lambda_2) > \partial t/\partial y(., \lambda_1)$ on (y'_c, y_c).

Let $y'_c < y_0 < y_c$, and define the increasing sequence $(y_n)_{n \in \mathbb{N}}$ by

$$
y_n = H^n_{12}(y_0)\mathbb{1}_{H_{12}(y_{n-1}) \in \mathcal{Y}_1} + y_{n-1}\mathbb{1}_{H_{12}(y_{n-1}) \notin \mathcal{Y}_1}.
$$

We see that $y_n < y_c$ for all $n \in \mathbb{N}$. Indeed, the result is true for $n = 0$. Moreover, if it holds for $n - 1$, then $y_n = H_{12}(y_{n-1}) < H_{12}(y_c) = y_c$ since H_{12} is strictly increasing. Hence, for all $n \in \mathbb{N}$, $y_n \in \mathcal{Y}_1$ and $y_n = H_{12}(y_{n-1})$. The sequence is increasing and bounded above by y_c, so it admits a limit y_∞ which satisfies $y_\infty = H_{12}(y_\infty)$. Hence $y_\infty = y_c$.

---

[21] At least locally around the true function for which assumption 1 is satisfied.

Now let θ_n be such that $\theta_n = \theta(y_n, \lambda_1)$. Note that because we fix θ_c and not θ_0, we cannot apply the proof of proposition 3.2 to show that θ_n is identified. Let us prove that θ_0 is identified.

First,

$$\theta_{n+1} = V_{21}(\theta_n, y_{n+1}) = \frac{\frac{\partial t}{\partial y}(y_{n+1}, \lambda_1)}{\frac{\partial t}{\partial y}(y_{n+1}, \lambda_2)} \theta_n.$$

Thus, by induction,

$$\theta_n = \prod_{i=1}^{n} \left[\frac{\frac{\partial t}{\partial y}(y_i, \lambda_1)}{\frac{\partial t}{\partial y}(y_i, \lambda_2)} \right] \theta_0.$$

Because $(y_n)_{n \in \mathbb{N}}$ converges to y_c and $\theta(., \lambda_1)$ is continuous, the sequence $(\theta_n)_{n \in \mathbb{N}}$ converges to $\theta_\infty = \theta_c$. Hence,

$$\theta_c = \prod_{i=1}^{\infty} \left[\frac{\frac{\partial t}{\partial y}(y_i, \lambda_1)}{\frac{\partial t}{\partial y}(y_i, \lambda_2)} \right] \theta_0.$$

Because $0 < \theta_c < \infty$, the product in the right hand side is strictly positive and finite and we have

$$\theta_0 = \frac{\theta_c}{\prod_{i=1}^{\infty} \left[\frac{\frac{\partial t}{\partial y}(y_i, \lambda_1)}{\frac{\partial t}{\partial y}(y_i, \lambda_2)} \right]}.$$

Hence, θ_0 is identified since the right term can be recovered from the data. y_0 was arbitrary, so $\theta(., \lambda_1)$ is identified on (y_c', y_c). Starting from θ_0 , we can also identify an increasing sequence $(\theta_n')_{n \in \mathbb{N}}$ which converges to θ_c' such that $y(\theta_c', \lambda_1) = y_c'$. This proves that $\theta(., \lambda_1)$ is actually identified on $[y_c', y_c]$. The same reasoning can be applied between y_c' and another crossing point y_c''. By repeating this as much as necessary, we can identify $\theta(., \lambda_1)$ on $\mathcal{Y}_1$.

Finally, by proposition 3.1, F_θ is identified on Θ and C' and $\widetilde{S}(., \lambda_1)$ are identified on $\mathcal{Y}_1$. Furthermore, by the horizontal transformation, $\theta(., \lambda_2)$ is also identified on $\mathcal{Y}_2$. Hence, $\widetilde{S}(., \lambda_2)$ is identified on $\mathcal{Y}_2$, and C' is identified on $\mathcal{Y}_1 \cup \mathcal{Y}_2$. ∎

Proof of proposition 3.4

$\widetilde{\mathbb{Y}}$ is the set of $\cup_{i=1,..,K} \mathcal{Y}_i$ such that the functions $\theta(., \lambda_i)$ are identified on $\widetilde{\mathbb{Y}} \cap \mathcal{Y}_i$. This set is defined by induction using the horizontal and vertical transformations.

Suppose that $y \in \widetilde{\mathbb{Y}} \cap \mathcal{Y}_i$ and that the point $(y, \theta(y, \lambda_i))$ is identified.

- For all j, $H_{ij}(y)$ and $\theta(H_{ij}(y), \lambda_j) = \theta(y, \lambda_i)$ are known because H_{ij} is identified. Hence the points $(H_{ij}(y), \theta(H_{ij}(y), \lambda_j))$ are identified.

- For all j such that $y \in \mathcal{Y}_j$, $\theta(y, \lambda_j) = V_{ij}(\theta(y, \lambda_i), y)$ is known because V_{ij} is identified. Hence the points $(y, \theta(y, \lambda_j))$ are identified if $y \in \mathcal{Y}_j$.

This method defines by induction $\widetilde{\mathbb{Y}}$, which corresponds to the set $\mathbb{Y}$ of the proposition. Furthermore, let $y \in \overline{\mathbb{Y}} \cap \mathcal{Y}_i$ and $y_n \in \mathbb{Y} \cap \mathcal{Y}_i$ such that $y_n \to y$. Then, by continuity of $\theta(., \lambda_i)$, we get

$$\theta(y, \lambda_i) = \lim_{n \to \infty} \theta(y_n, \lambda_i).$$

$\theta(., \lambda_i)$ is actually identified on $\overline{\mathbb{Y}} \cap \mathcal{Y}_i$. ∎

Proof of theorem 3.5

Let us define $\widetilde{\Theta} = \theta(\mathbb{Y} \cap \mathcal{Y}_1, \lambda_1)$, $E_1 = l(\lambda_1)/l(\lambda_2)$, $E_3 = l(\lambda_3)/l(\lambda_2)$ and $\Theta = [\underline{\theta}, \overline{\theta}]$. The result is based on the following two lemmas.

Lemma 6.1 *Suppose that*

$$\{E_1^{m_1} E_3^{m_3} \theta_0, \ (m_1, m_3) \in \mathbb{Z}^2\} \cap \Theta \subset \widetilde{\Theta}. \tag{6.6}$$

Then $\overline{\widetilde{\Theta}} = \Theta$.

Lemma 6.2 *If* $\overline{\widetilde{\Theta}} = \Theta$, *then* $\theta(., \lambda_1)$ *is identified on* $\mathcal{Y}_1$.

Proof of lemma 6.1: let us introduce

$$G = \{m_1 \ln(E_1) + m_3 \ln(E_3), (m_1, m_3) \in \mathbb{Z}^2\}.$$

G is an additive subgroup of $\mathbb{R}$. Thus, it is either of the form $a\mathbb{Z}$ or dense in $\mathbb{R}$. By assumption 5, $\ln(E_1)/\ln(E_3)$ is irrational. Thus G is dense in $\mathbb{R}$. By continuity of the exponential, $\{E_1^{m_1} E_3^{m_3} \theta_0, \ (m_1, m_3) \in \mathbb{Z}^2\}$ is dense in $\mathbb{R}^+$. Hence, by (6.6), $\Theta \subset \overline{\widetilde{\Theta}}$. The other inclusion is obvious, so $\overline{\widetilde{\Theta}} = \Theta$. □

Proof of lemma 6.2: by continuity of $y(., \lambda_1)$, the inverse of $\theta(., \lambda_1)$, we get

$$y(\overline{\widetilde{\Theta}}, \lambda_1) \subset \overline{y(\widetilde{\Theta}, \lambda_1)} = \overline{\mathbb{Y}} \cap \mathcal{Y}_1.$$

Thus, if $\overline{\widetilde{\Theta}} = \Theta$, $\mathcal{Y}_1 = y(\overline{\widetilde{\Theta}}, \lambda_1) \subset \overline{\mathbb{Y}}$ and by proposition 3.4, $\theta(., \lambda_1)$ is identified on $\mathcal{Y}_1$. □

Now, let us come back to the proof of the theorem. By assumption 6, there exists $y_0 \in \mathcal{Y}_1 \cap \mathcal{Y}_3$ such that $H_{23}(y_0) \in \mathcal{Y}_2$ and $H_{21}(y_0) \in \mathcal{Y}_2$. Let θ_0 satisfy $\theta(y_0, \lambda_2) = \theta_0$. By assumption 4, the vertical transforms write as $V_{2i}(\theta, y) = E_i \theta$ (for $i = 1, 3$) provided that

$y \in \mathcal{Y}_2 \cap \mathcal{Y}_i$. Moreover, by definition of y_0, $\theta(y_0, \lambda_i) = E_i \theta_0$ is well defined (i.e., $E_i \theta_0 \in \Theta$). Hence,

$$\theta(H_{31}(y_0), \lambda_1) = E_3 \theta_0.$$

In other terms, $E_3 \theta_0 \in \widetilde{\Theta}$. Similarly, $E_1 \theta_0 \in \widetilde{\Theta}$.

Moreover, $H_{23}(y_0) \in \mathcal{Y}_2 \cap \mathcal{Y}_3$, so that $V_{32}(\theta_0, H_{23}(y_0)) = E_3^{-1} \theta_0$ is well defined. Hence,

$$\theta(H_{21}(H_{23}(y_0)), \lambda_1) = E_3^{-1} \theta_0.$$

Thus, $E_3^{-1} \theta_0 \in \widetilde{\Theta}$ (and similarly for $E_1^{-1} \theta_0$).

Now, for all $n \geq 1$, define

$$A_n = \{E_1^{m_1} E_3^{m_3} \theta_0, \; |m_1| + |m_3| = n\} \cap \Theta$$

Let us show by induction on n that $A_n \subset \widetilde{\Theta}$. The result is true for $n = 1$ by the preceding. Suppose that it is true for n and let $\theta = E_1^{m_1} E_3^{m_3} \theta_0 \in A_{n+1}$, with $\theta \geq \theta_0$ (the case where $\theta < \theta_0$ is similar). We have to show that $\theta \in \widetilde{\Theta}$.

Suppose first that $m_1 < 0$. Because $\theta > E_1 \theta \geq E_1 \theta_0 \geq \underline{\theta}$, $E_1 \theta \in \Theta$. Moreover,

$$E_1 \theta = E_1^{m_1+1} E_3^{m_3} \theta_0, \quad |m_1 + 1| + |m_3| = (-m_1 - 1) + |m_3| = n.$$

Thus $E_1 \theta \in A_n \subset \widetilde{\Theta}$ by the induction hypothesis. By definition of $\widetilde{\Theta}$, there exists $y \in \mathbb{Y}$ such that $\theta(y, \lambda_1) = E_1 \theta$. $\theta = V_{12}(E_1 \theta, y)$ is well defined because $\theta \in \Theta$. This implies that $H_{21}(y) \in \mathbb{Y} \cap \mathcal{Y}_1$. Moreover, $\theta(H_{21}(y), \lambda_1) = \theta$. Thus, $\theta \in \widetilde{\Theta}$.

Similarly, if $m_1 \geq 0$ and $m_3 > 0$, the same reasoning applies, using $E_3^{-1} \theta$ instead of $E_1 \theta$.[22] $A_{n+1} \subset \widetilde{\Theta}$ and the result is true for all n. Hence, $\cup_n A_n \subset \widetilde{\Theta}$, so that (6.6) holds.

Finally, by lemmas 6.1 and 6.2, $\theta(., \lambda_1)$ is identified on $\mathcal{Y}_1$. We conclude as in theorem 3.3. ∎

Proof of proposition 3.6

By equations (3.1) and (3.2), $\theta(., \lambda)$ and $\partial C / \partial y(., .)$ are identified on respectively $\mathcal{Y}_\lambda = \{y / \exists \theta \in \Theta / \theta(y, \lambda) = \theta\}$ and $\{(y, \theta(y, \lambda)), y \in \mathcal{Y}_\lambda\}$, for all $\lambda \in \Lambda$. Now, for all $y \in \mathcal{Y}_\lambda$,

$$\frac{\partial^2 C}{\partial y \partial \lambda}(y, \theta(y, \lambda)) = \frac{\partial^2 C}{\partial y \partial \theta}(y, \theta(y, \lambda)) \frac{\partial \theta}{\partial \lambda}(y, \lambda).$$

---

[22] The case where $m_1 \geq 0$ and $m_3 \leq 0$ is impossible, since we assumed that $\theta \geq \theta_0$.

Moreover, $\partial^2 t/\partial y \partial\lambda(y,\lambda) > 0$ implies that $\partial\theta/\partial\lambda(y,\lambda) > 0$. Thus, for all $y \in \mathcal{Y}_\lambda$, $\partial^2 C/\partial y \partial\theta(y, \theta(y,\lambda))$ is identified by

$$\frac{\partial^2 C}{\partial y \partial\theta}(y, \theta(y,\lambda)) = \frac{\partial^2 C/\partial y \partial\lambda(y, \theta(y,\lambda))}{\partial\theta/\partial\lambda(y,\lambda)}.$$

By equation (3.3), $\widetilde{S}(.,\lambda)$ is therefore identified on $\mathcal{Y}_\lambda$. ■

Proof of proposition 3.7

When $K = 1$ and C', F_θ and $\widetilde{S}$ are unknown, neither model implies any testable restriction. If one of these three functions is known, proposition 3.1 shows that the asymmetric model is just identified. Using equations (2.1), (2.2) and equation (6.2) in appendix A, it can be shown that the complete information model is also identified. In general, no contradiction arises in either model, except if assumption 1 fails to hold in one model but not in the other.

When $K \geq 2$, in the complete information model, equations (6.2) and (2.2) show that

$$t(y, \lambda_1) > t(y, \lambda_2) \iff F_y(y, \lambda_1) < F_y(y, \lambda_2).$$

This condition is slightly different from (3.10) and we are able to distinguish the two models if (3.11) holds.

Otherwise, the analysis is more involved. The identification of the complete information model with exogenous variations follows the same lines than previously. Using equations (2.1) and (2.2), we recover $\widetilde{S}(.,\lambda)$ and $C(.)$ from $\theta(., \lambda_1)$. By the separability hypothesis, this function is in turn identified through the same horizontal and vertical transforms as in the incomplete model. Hence, using an asymmetric information model when the true model is the complete information one (or the contrary) actually leads to the same $\theta(., \lambda_1)$, and consequently to the same F_θ.

Differentiating the first order condition in the symmetric case leads to

$$t'(y) = \theta(y)C^{S\prime}(y) + \theta'(y)C^S(y).$$

On the other hand, C^A satisfies

$$t'(y) = \theta(y)C^{A\prime}(y).$$

Because $\theta'(y)C^S(y) < 0$, $C^{A\prime} < C^{S\prime}$ and the marginal cost functions differ. Hence, if C' is known, the models are distinguishable.

Now let us suppose that $\widetilde{S}(., \lambda_0)$ is known. The functions $\widetilde{S}^S(., \lambda_0)$ and $\widetilde{S}^A(., \lambda_0)$ obtained respectively with the symmetric and asymmetric model satisfy

$$\widetilde{S}^S(y, \lambda_0) = \theta(y, \lambda_0)C^{S\prime}(y) = t'(y, \lambda_0) - t(y, \lambda_0)\frac{\theta'(y, \lambda_0)}{\theta(y, \lambda_0)}.$$

$$\widetilde{S}^A(y, \lambda_0) = \left[1 - \frac{1 - F_y(y, \lambda_0)}{f_y(y, \lambda_0)}\frac{\theta'(y, \lambda_0)}{\theta(y, \lambda_0)}\right] t'(y, \lambda_0).$$

Thus, $\widetilde{S}^S(., \lambda_0)$ and $\widetilde{S}^A(., \lambda_0)$ are identical if and only if, for all y,

$$\frac{t'(y, \lambda_0)}{t(y, \lambda_0)} = \frac{f_y(y, \lambda_0)}{1 - F_y(y, \lambda_0)}$$

Integrating this expression leads to

$$t(y, \lambda_0) = \frac{K}{1 - F_y(y, \lambda_0)}.$$

for a given $K > 0$. Thus, we can distinguish the models if and only if $y \mapsto t(y, \lambda_0) \times (1 - F_y(y, \lambda_0))$ is not constant. $\blacksquare$

Proof of theorem 4.1

In the following, we denote $F_j = F(., \lambda_j)$ and $\theta_j = \theta(., \lambda_j)$, $(j \in \{1, 2\})$. The result is based on the six following lemmas.

Lemma 6.3 *Let $j \in \{1, 2\}$ and K denote a compact set in $\overset{\circ}{\mathcal{Y}}_j$. Then*

$$\inf_{y \in K} f_j(y) > 0.$$

Lemma 6.4 *Let $j \in \{1, 2\}$ and K denote a compact set in $(0, 1)$. Then*

$$\sup_{y \in K} |\widehat{F}_j^{-1}(y) - F_j^{-1}(y)| \overset{\mathbb{P}}{\longrightarrow} 0.$$

Lemma 6.5 *For all compact set $K \subset \overset{\circ}{\mathcal{Y}}_1$,*

$$\sup_{x \in K} |\widehat{H}_{12}(x) - H_{12}(x)| \overset{\mathbb{P}}{\longrightarrow} 0.$$

Lemma 6.6 *If $\widehat{y}_{n-1}$ converges in probability to y_{n-1}, then*

$$\widehat{H}_{12}(\widehat{y}_{n-1}) \overset{\mathbb{P}}{\longrightarrow} H_{12}(y_{n-1}).$$

Lemma 6.7 *If $\widehat{y}_{n-1}$ converges in probability to y_{n-1}, then*

$$\mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1})\in\widehat{\mathcal{Y}_1}} \xrightarrow{\mathbb{P}} \mathbb{1}_{H_{12}(y_{n-1})\in\mathcal{Y}_1}.$$

$$\mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1})\notin\widehat{\mathcal{Y}_1}} \xrightarrow{\mathbb{P}} \mathbb{1}_{H_{12}(y_{n-1})\notin\mathcal{Y}_1}.$$

Lemma 6.8 *For all $n \in \mathbb{Z}$ and when $N \to +\infty$,*

$$(\widehat{\theta}_n, \widehat{y}_n) \xrightarrow{\mathbb{P}} (\theta_n, y_n).$$

Proof of lemma 6.3: let (for instance) $j = 1$ and let $\theta_1 \equiv \theta(., \lambda_1)$. By equation (3.6), we get for all $y \in \overset{\circ}{\mathcal{Y}_1}$,

$$f_1(y) = -f_\theta(\theta_1(y))\theta_1'(y).$$

Now, deriving the first order condition, we get, because $t'(., \lambda_1) = \delta_1$,

$$\theta_1'(y) = -\frac{\delta_1}{C'^2(y)}C''(y).$$

Thus, by assumption 1, $\theta_1'(y) < 0$ for all $y \in \overset{\circ}{\mathcal{Y}_1}$. Moreover, because K is a compact subset strictly included in $\mathcal{Y}_1$, $f_\theta(\theta_1(y)) > 0$ for all $y \subset K$. Hence, by continuity of f_θ and θ',

$$\inf_{y\in K} f_1(y) = \min_{y\in K}\left[-f_\theta(\theta_1(y))\theta_1'(y)\right] > 0 \;\square$$

Proof of lemma 6.4: let $j = 1$ and $\varepsilon > 0$ be such that $E = \{x \in \mathbb{R}/\exists y \in F_1^{-1}(K)/|x-y| \leq \varepsilon\}$ is a subset of $\overset{\circ}{\mathcal{Y}_1}$. E is compact, so by the previous lemma, $m = \inf_{x\in E} f_1(x) > 0$. Moreover, for all $y \in F_1^{-1}(K)$, $F_1(y - \varepsilon) + m\varepsilon \leq F_1(y) \leq F_1(y + \varepsilon) - m\varepsilon$. Consequently,

$$\sup_{y\in K}|\widehat{F}_1^{-1}(y) - F_1^{-1}(y)| > \varepsilon \iff \exists y \in K \ / \ |\widehat{F}_1^{-1}(y) - F_1^{-1}(y)| > \varepsilon$$

$$\iff \exists y \in K \ / \ F_1(F_1^{-1}(y)) > \widehat{F}_1(F_1^{-1}(y) + \varepsilon)$$
$$\text{or } F_1(F_1^{-1}(y)) < \widehat{F}_1(F_1^{-1}(y) - \varepsilon)$$
$$\implies \exists y \in K \ / \ F_1(F_1^{-1}(y) + \varepsilon) - \widehat{F}_1(F_1^{-1}(y) + \varepsilon) > m\varepsilon$$
$$\text{or } \widehat{F}_1(F_1^{-1}(y) - \varepsilon) - F_1(F_1^{-1}(y) - \varepsilon) > m\varepsilon$$
$$\implies \sup_{x\in\mathcal{Y}_1}|\widehat{F}_1(x) - F_1(x)| > m\varepsilon$$

Hence,

$$\mathbb{P}(\sup_{y\in K}|\widehat{F}_1^{-1}(y) - F_1^{-1}(y)| > \varepsilon) \leq \mathbb{P}(\sup_{x\in\mathcal{Y}_1}|\widehat{F}_1(x) - F_1(x)| > m\varepsilon).$$

The right term tends to zero by the Glivenko-Cantelli theorem. Thus, $\sup_{y\in K}|\widehat{F}_1^{-1}(y) - F_1^{-1}(y)|$ converges to zero in probability. $\square$

Proof of lemma 6.5: by the triangular inequality,

$$\sup_{x\in K}|\widehat{H}_{12}(x) - H_{12}(x)| \leq \sup_{x\in K}|\widehat{F}_2^{-1}\circ\widehat{F}_1(x) - F_2^{-1}\circ\widehat{F}_1(x)| + \sup_{x\in K}|F_2^{-1}\circ\widehat{F}_1(x) - F_2^{-1}\circ F_1(x)|$$
(6.7)

Let us show that the two terms in the right hand side converge to zero. Because $K = [\underline{K},\overline{K}]\subset\overset{\circ}{\mathcal{Y}}_1$, there exists a compact set $L\subset(0,1)$ such that $F_1(K)\subset L$. Thus, for all $\varepsilon > 0$, since $\widehat{F}_1$ is increasing,

$$\mathbb{P}(\sup_{x\in K}|\widehat{F}_2^{-1}\circ\widehat{F}_1(x) - F_2^{-1}\circ\widehat{F}_1(x)| > \varepsilon) \leq \mathbb{P}(\widehat{F}_1(\underline{K})\notin L\cup\widehat{F}_1(\overline{K})\notin L)$$
$$+ \mathbb{P}(\sup_{y\in L}|\widehat{F}_2^{-1}(y) - F_2^{-1}(y)| > \varepsilon).$$

The first term of the right hand side converges to zero since $\widehat{F}_1(\underline{K})$ (resp. $\widehat{F}_1(\overline{K})$) converges in probability to $F_1(\underline{K})\in L$ (resp. $F_1(\overline{K})\in L$). The second term tends to zero by the previous lemma. Thus the first term of (6.7) converges to zero in probability.

Let us turn to the second term of (6.7). F_2^{-1} is uniformly continuous on $F_1(K)$, as a continuous function on a compact. Thus, there exists η such that $|x - y|\leq\eta$ implies $|F_2^{-1}(x) - F_2^{-1}(y)|\leq\varepsilon$. Hence,

$$\mathbb{P}(\sup_{x\in K}|F_2^{-1}\circ\widehat{F}_1(x) - F_2^{-1}\circ F_1(x)| > \varepsilon) \leq \mathbb{P}(\sup_{x\in K}|\widehat{F}_1(x) - F_1(x)| > \eta).$$

The right term converges to zero by the Glivenko-Cantelli theorem, implying the result. $\square$

Proof of lemma 6.6: For all $A,\varepsilon > 0$,

$$\mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > A) \leq \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > A, |\widehat{y}_{n-1} - y_{n-1}|\leq\varepsilon)$$
$$+ \mathbb{P}(|\widehat{y}_{n-1} - y_{n-1}| > \varepsilon).$$

Let K denote a compact set in $\overset{\circ}{\mathcal{Y}}_1$ such that $y_{n-1}\in\overset{\circ}{K}$. Because H'_{12} is continuous on K, $\max_{x\in K}|H'_{12}(x)| = M$ exists. Let $\varepsilon > 0$ be such that $M\varepsilon < A$ and the closed ball of radius ε centered at y_{n-1} is a subset of K. Then, when $|\widehat{y}_{n-1} - y_{n-1}|\leq\varepsilon$,

$$|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \leq |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(\widehat{y}_{n-1})| + |H_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})|$$
$$< \sup_{x\in K}|\widehat{H}_{12}(x) - H_{12}(x)| + M\varepsilon.$$

Thus,

$$\mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})|\geq A) \leq \mathbb{P}\left(\sup_{x\in K}|\widehat{H}_{12}(x) - H_{12}(x)| > A - M\varepsilon\right)$$
$$+ \mathbb{P}(|\widehat{y}_{n-1} - y_{n-1}| > \varepsilon)$$

By assumption, the second term tends to zero in probability. $\sup_{x \in K} |\widehat{H}_{12}(x) - H_{12}(x)|$ also converges to zero by lemma 6.5. The result follows. $\square$

Proof of lemma 6.7: We consider two cases. First, if $H_{12}(y_{n-1}) \notin \mathcal{Y}_1$, then, because $\mathcal{Y}_1$ is a closed set, $\min_{x \in \mathcal{Y}_1} |H_{12}(y_{n-1}) - x| = a > 0$. Now, because $\widehat{\mathcal{Y}_1} \subset \mathcal{Y}_1$,

$$
\begin{aligned}
\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}_1}) &\leq \mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \in \mathcal{Y}_1) \\
&\leq \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \geq a)
\end{aligned}
$$

By lemma 6.6, $\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}_1})$ tends to zero.

Secondly, suppose that $H_{12}(y_{n-1}) \in \mathcal{Y}_1$. Then, by assumption 8 and because $H_{12}(y_{n-1}) = y_n$, $H_{12}(y_{n-1}) \in \overset{\circ}{\mathcal{Y}_1}$ and $0 < F_1(H_{12}(y_{n-1})) < 1$. Hence, there exists $a > 0$ such that $F_1(H_{12}(y_{n-1}) + a) < 1$ and $F_1(H_{12}(y_{n-1}) - a) > 0$. Now,

$$
\begin{aligned}
\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1}) &\leq \mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1}, |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \leq a) \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq \mathbb{P}\Big[\{\widehat{H}_{12}(\widehat{y}_{n-1}) > \max_k y_{1k}\} \cup \{\widehat{H}_{12}(\widehat{y}_{n-1}) < \min_k y_{1k}\}, \\
&\qquad |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \leq a \Big] + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq \mathbb{P}\Big[\{H_{12}(y_{n-1}) + a > \max_k y_{1k}\} \cup \{H_{12}(y_{n-1}) - a < \min_k y_{1k}\} \Big] \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq \mathbb{P}(H_{12}(y_{n-1}) + a > \max_k y_{1k}) + \mathbb{P}(H_{12}(y_{n-1}) - a < \min_k y_{1k}) \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq [F_1(H_{12}(y_{n-1}) + a)]^N + [1 - F_1(H_{12}(y_{n-1}) - a)]^N \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a).
\end{aligned}
$$

The first two terms converge to zero. The third term tends to zero in probability by lemma 6.6. Thus $\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1})$ tends to zero. This proves the lemma. $\square$

Proof of lemma 6.8: We proceed by induction. The result holds for $n = 0$ since $\widehat{y}_0 = y_0$ and $\widehat{\theta}_0 = \theta_0$. Suppose that it is true for $n - 1 \geq 0$. Using the induction hypothesis, lemma 6.6 and 6.7, $\widehat{y}_n$ converges to y_n in probability. Similarly, using lemma 6.7, $\widehat{\theta}_n$ converges to θ_n. Because convergence in probability of each term implies joint convergence (see e.g. Van der Vaart, 1998, theorem 2.7), the result holds for n. Hence, lemma 6.8 is true for all $n \geq 0$. The proof is similar for negative values. $\square$

Now, let us come back to the proof of theorem 4.1. By the previous lemma, it suffices to prove that $\widehat{F_\theta(\theta_n)}$ (resp. $\widehat{C'(y_n)}$) converges in probability to $F_\theta(\theta_n)$ (resp. $C'(y_n)$). By the

triangular inequality,

$$
\begin{aligned}
|\widehat{F_\theta(\theta_n)} - F_\theta(\theta_n)| &= |\widehat{F}_1(\widehat{y}_n) - F_1(y_n)| \\
&\leq |\widehat{F}_1(\widehat{y}_n) - F_1(\widehat{y}_n)| + |F_1(\widehat{y}_n) - F_1(y_n)| \\
&\leq \sup_{y \in \mathcal{Y}_1} |\widehat{F}_1(y) - F_1(y)| + |F_1(\widehat{y}_n) - F_1(y_n)|.
\end{aligned}
$$

The first term converges to zero by the Glivenko-Cantelli theorem. The second term also converges to zero by the previous lemma and the continuity of F_1. The first convergence follows. Because $\widehat{C'(y_n)} = \delta_1/\widehat{\theta}_n$ is a continuous function of $\widehat{\theta}_n$, the second result stems from lemma 6.8 once more. ∎

7 Appendix B: surplus

Estimation of the parameters a, b, α, β

The first order condition of the agent writes as $\theta(y, \lambda_i) = \delta_i/C'(y)$. Using the parametric form of the cost $C'(y) = \alpha \left(\frac{y}{1-y} \right)^\beta$, we obtain

$$
\theta(y, \lambda_i) = \frac{\delta_i}{\alpha} \left(\frac{1-y}{y} \right)^\beta
$$

Then, using the Weibull specification, we obtain

$$
F_y(y, \lambda_i) = 1 - F_\theta(\theta(y, \lambda_i)) = \exp\left(-a \left(\frac{\delta_i}{\alpha} \right)^b \left(\frac{1-y}{y} \right)^{b\beta} \right)
$$

which rewrites as

$$
\ln(-\ln(F_y(y, \lambda_i))) = \ln(a) + b\left(\ln(\delta_i) - \ln(\alpha) \right) + b\beta \ln\left(\frac{1-y}{y} \right).
$$

The parameters are estimated by regressing $\ln(-\ln(\widehat{F}_j(y)))$ on $\ln(\frac{1-y}{y})$. As explained in the main text, a normalization is necessary and we impose $C'(y_0) = \delta_1/\theta_0$.

Estimation of λ_3

We suppose here that Insee can only implement linear contracts $t(y, \delta) = \delta y$ and denote by $y(\theta, \delta)$ the response rate chosen by a θ type agent. Insee solves, for the 2003 survey,

$$
\max_\delta \int (\lambda_3 - \delta) y(\theta, \delta) f_\theta(\theta) d\theta.
$$

and the first order condition writes as

$$
-\int y(\theta, \delta_3) f_\theta(\theta) d\theta + (\lambda_3 - \delta_3) \int \frac{\partial y}{\partial \delta}(\theta, \delta_3) f_\theta(\theta) d\theta = 0.
$$

Using again the first order condition of the agent, we have $y(\theta, \delta) = \frac{1}{1+\left(\frac{\alpha\theta}{\delta}\right)^{1/\beta}}$. Thus,

$$\frac{\partial y}{\partial \delta}(\theta, \delta) = \frac{1}{\beta\delta}y(1-y).$$

and the condition takes the form

$$-E(y(\theta, \delta_3)) + \frac{\lambda_3 - \delta_3}{\beta\delta_3}E(y(\theta, \delta_3)(1 - y(\theta, \delta_3))) = 0$$

Hence,

$$\lambda_3 = \delta_3\left(1 + \beta\frac{E(y(\theta, \delta_3))}{E(y(\theta, \delta_3)(1 - y(\theta, \delta_3)))}\right)$$

which is estimated, noting $\bar{y}_3$ (resp. $\overline{y^2}_3$) the empirical mean of the response rate (resp. the empirical mean of the response rate's square) in 2003, by

$$\widehat{\lambda}_3 = \delta_3\left(1 + \widehat{\beta}\frac{\bar{y}_3}{\bar{y}_3 - \overline{y^2}_3}\right)$$

Linear contracts under incomplete information

The expected surplus of Insee, when linear contracts are used, is $\Pi = (\lambda_3 - \delta_3)E(y(\theta, \delta_3))$ which can be estimated by

$$\widehat{\Pi} = (\widehat{\lambda}_3 - \widehat{\delta}_3)\bar{y}_3$$

Optimal contracts under incomplete information

On one hand, under incomplete information, the optimal contract is defined by

$$\lambda_3 = \left[\theta + \frac{F_\theta(\theta)}{f_\theta(\theta)}\right]C'(y^I(\theta, \lambda_3)).$$

Hence, using the form of the cost function, we obtain

$$y^I(\theta, \lambda_3) = \frac{1}{1 + \left(\frac{\alpha}{\lambda_3}\right)^{1/\beta}\left[\theta + \frac{F_\theta(\theta)}{f_\theta(\theta)}\right]^{1/\beta}}.$$

On the other hand, using the first order condition of the agent, we have

$$t'(y^I(\theta, \lambda_3)) = \theta C'(y^I(\theta, \lambda_3)) = \lambda_3\frac{\theta}{\theta + F_\theta(\theta)/f_\theta(\theta)}.$$

Hence, the surplus under incomplete information, which writes as

41

$$\Pi^I = \lambda_3 E(y^I(\theta, \lambda_3)) - E(t(y^I(\theta, \lambda_3)))$$

can be estimated by Monte-Carlo simulations.

More precisely, to estimate this surplus, we draw 100,000 values of θ in our estimated Weibull distribution $F_\theta(.|\widehat{a}, \widehat{b})$ and compute the previous functions using the parameters $\widehat{\lambda}_3$, $\widehat{\alpha}$ and $\widehat{\beta}$.

Optimal contracts under complete information

Under complete information, the first order condition for the agent writes $\lambda_3 = \theta C'(y^C(\theta, \lambda_3))$. Hence, using the parametric form of the cost function, we obtain

$$y^C(\theta, \lambda_3) = \frac{1}{1 + \left(\frac{\alpha\theta}{\lambda_3}\right)^{1/\beta}}.$$

Furthermore, the transfer function is given by

$$t(y^C(\theta, \lambda_3), \lambda_3) = \theta C(y^C(\theta, \lambda_3)) = \alpha\theta \int_0^{y^C(\theta, \lambda_3)} \left(\frac{u}{1-u}\right)^\beta du$$

Finally, the expected surplus under complete information is given by

$$\Pi^C = \lambda_3 E(y^C(\theta, \lambda_3)) - E(t(y^C(\theta, \lambda_3))).$$

As previously, to estimate this surplus, we draw 100,000 values of θ in our estimated Weibull distribution $F_\theta(.|\widehat{a}, \widehat{b})$ and compute the previous functions using the parameters $\widehat{\lambda}_3$, $\widehat{\alpha}$ and $\widehat{\beta}$.

References

Abbring J. H., J. J. Heckman, P. A. Chiappori and J. Pinquet (2003), Adverse Selection and Moral Hazard In Insurance: Can Dynamic Data Help to Distinguish?, Journal the European Economic Association 1, 512–521.

Akerlof, G. (1970), The Market for "Lemons": Quality Uncertainty and the Market Mechanism, The Quarterly Journal of Economics 84, 488–500.

Ausubel, L. M. (1999), Adverse Selection in the Credit Card Market, Mimeo, University of Maryland.

Baron D. P. and R. B. Myerson (1982), Regulating a Monopolist with Unknown Costs, Econometrica 4, 911–930.

Banerjee, A. P. Gertler and M. Ghatak (2002), Empowerment and Efficiency: Tenancy Reform in West Bengal, Journal of Political Economy 110, 239–280.

Chiappori, P. A., F. Durand and P.Y. Geoffard (1998), Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment, European Economic Review 42, 499–511.

Chiappori, P. A., P.Y. Geoffard and E. Kyriadizou (1998), Cost of Time, Moral Hazard, and the Demand for Physician Services, Mimeo, University of Chicago.

Chiappori, P. A., B. Jullien, B. Salanié and F. Salanié (2006), Asymmetric Information in Insurance: General Testable Implications, The RAND Journal of Economics.

Chiappori, P. A. and B. Salanié (2000), Testing for Asymmetric Information in Insurance Markets, Journal of Political Economy 108, 56–78.

Chiappori, P. A. and B. Salanié (2002), Testing Contract Theory: A Survey of Some Recent Work, in Advances in Economics and Econometrics, vol 1, M. Dewatripont, L. Hansen et S. Turnovsky eds, Cambridge University Press.

Diamond, P. (1998), Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates, The American Economic Review 88, 83–95.

Dionne, G. and C. Vanasse (1996), Une évaluation empirique de la nouvelle tarification de l'assurance automobile au Québec, Actualité économique 73, 47–80.

Ferrall C. and B. Shearer (1999), Incentives and Transaction costs within the firm: Estimating an Agency Model Using Payroll Records, Review of Economic Studies 66, 309–338.

Freixas X. and J. J. Laffont (1990), Optimal Banking Contracts, Essays in Honor of Edmond Malinvaud, vol. 2: Macroeconomics, MIT Press, 33–61.

Freixas X. and J. C. Rochet (1997), The Microeconomics of Banking, Microeconomics of Banking, MIT Press.

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.

Gagnepain P. and M. Ivaldi (2007), Contract Choice, Incentives, and Political Capture in Public Transit, mimeo.

Gence-Creux C. (2000), Regulation with a Risk-Averse Principal, in: Econometric Society World Congress 2000 Contributed Papers.

Guerrre E., I. Perrigne, and Q. Vuong (2005), Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions, Mimeo.

Haubrich J. (1994) Risk Aversion, Performance Pay, and the Principal-Agent Model, Journal of Political Economy 102, 258–276.

Ivaldi M., and D. Martimort (1994), Competition under Nonlinear Pricing, Annales d'Economie et de Statistique 34, 71–114.

Jensen M. and K. Murphy (1990), Performance Pay and Top-Management Incentives, Journal of Political Economy 98, 225–264.

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

Laffont, J. J. and J. Tirole (1986), Using Cost Observation to Regulate Firms, The Journal of Political Economy 94, 614–641.

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

Lavergne P., and A. Thomas (2005), Semiparametric estimation and testing in a model of environmental regulation with adverse selection, Empirical Economics 30, 171–192.

Lazear, E. (2000), Performance Pay and Productivity, American Economic Review 90, 1346–1361.

Lewis, T. and D. Sapppington (1995), Optimal Capital Structure in Agency Relationships, The RAND Journal of Economics 26, 343–361.

Manning, W., J. Newhouse, N. Duan, E. Keeler and A. Leibowitz (1987), Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment, American Economic Review 77, 251–277.

Maskin, E. and J. Riley (1984), Monopoly with Incomplete Information, The RAND Journal of Economics 15, 171–196.

Miravette E. J. (2002), Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans, Review of Economic Studies 69, 943–971.

Mussa M. and S. Rosen (1978), Monopoly and product quality, Journal of Economic Theory, 18, 301–317.

Myerson R. B. (1981), Optimal Auction Design, Mathematics of Operation Research 6, 58–73.

Perrigne, I. (2002), Incentive Regulatory Contracts in Public Transportation: An Empirical Study, Working Paper, Pennsylvania State University.

Perrigne, I and Q. Vuong (2004), Nonparametric Identification of Incentive Regulation Models, Mimeo.

Prendergast, C. (1999), The Provision of Incentives in Firms, Journal of Economic Literature 37, 7–63.

Puelz, R. and A. Snow (1994), Evidence on Adverse Selection: Equilibrium Signalling and Cross-Subsidization in the Insurance Market, Journal of Political Economy, 102, 236–57.

Rosen S. (1992), Contracts and The Market for Executives, in Contract Economics, L. Werin and H. Wijkander eds, Basil Blackwell.

Salanié, B. (2005a), The Economics of Contracts: a Primer, 2nd edition, MIT Press.

Salanié, F. (2005b), On screening, Working Paper.

Shearer, B. (2004), Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment, Review of Economic Studies 71, 513–534.

Wilson, R. (1993), Nonlinear pricing, Oxford University Press.

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