Inference on Adverse Selection Models through Contracts Variation *

Xavier D'Haultfoeuille † Philippe Février‡

May 2009

Abstract

We study the empirical content of an adverse selection model defined by the objective function of the principal, the agents' cost function and the distribution of agents' types. This model, though simple, encompasses the basic regulation, nonlinear pricing and price discrimination models, first price auctions and simple insurance settings. We prove that nonparametric identification of the model requires the knowledge of at least one of the three functions. We also show that one or two 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. Finally, we apply this method to contract data between the French National Institute of Statistics and its interviewers. We estimate the loss of using two part tariffs 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

*This manuscript is a substantial revision of the working paper "Identification and Estimation of Incentive Problems: Adverse Selection". We are very grateful to three anonymous referees, Steve Berry, Thierry Magnac, Martin Pesendorfer, Rob Porter, Jean-Marc Robin, Jean-Charles Rochet, Mathieu Rosenbaum, Bernard Salanié, and the participants of various seminars and conferences for helpful discussions and comments. We also thank Daniel Verger and the Unité Méthodes Statistiques of Insee for providing us with the data.

†ENSAE, CREST-INSEE and Université Paris I-Panthéon-Sorbonne. E-mail address: xavier.dhaultfoeuille@ensae.fr.

‡CREST-INSEE (Paris). E-mail address: fevrier@ensae.fr.

1 Introduction

Over the past three decades, 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 has been helpful, for instance, to better understand theoretically nonlinear pricing, regulation, financial contracts or taxation theory. In recent years, the empirical literature on adverse selection models has also grown rapidly.[1] However, apart from the auction literature, most of these papers adopt a parametric framework.[2] Such parametric restrictions obscure what can be identified nonparametrically from the model and data. This issue is yet important to safely investigate optimality of contracts or do policy exercises, without being sensitive to a particular choice of parametrization.

In this paper, we analyze the nonparametric inference on a simple adverse selection model when the econometrician observes the contract and the associated trades. The model we consider is characterized by the objective function of the principal, the distribution of agents' types and the cost function of the agents. The space of trades available to the agent is supposed to be an interval of the real line, the agent's type is unidimensional and the cost is assumed to be separable. Such conditions discard discrete choices, which are common in price discrimination for instance, as well as multidimensional screening, and reduce the dimensionality of the cost function. Despite these limitations, the model encompasses the basic regulation, nonlinear pricing and price discrimination models, financial contracts, delegation of tasks by firms, first price auctions or simple insurance settings. Thus, even if our model cannot take the particular features of each possible application into account, we believe that our results deliver useful insights on the empirical content of adverse selection models, in a spirit close to what has been done in auctions, building on the work of Guerre et al. (2000). Understanding the econometrics of the most simple common structure of

[1] Applications include auction models (see e.g. Paarsch, 1992 and Guerre et al., 2000), regulatory contracts (see, among others, Wolak, 1994, Gagnepain and Ivaldi, 2002, Perrigne, 2002, Perrigne and Vuong, 2004 and Lavergne and Thomas, 2005), nonlinear pricing (see Ivaldi and Martimort, 1994, Miravete, 2002, Miravete and Roller, 2005, Huang et al., 2007 and Miravete, 2007) and price discrimination models (see Leslie, 2004 and Crawford and Shum, 2007). Adverse selection models have also been used to study the provision of incentives in firms, see e.g. Ferrall and Shearer (1999) and Paarsch and Shearer (2000).

[2] Notable exceptions are the papers of Perrigne and Vuong (2004) and Huang et al. (2007). The first studies the Laffont and Tirole (1986) regulation model in which ex-post costs are observed, and shows that such a model is nonparametrically identified. The second considers a nonlinear pricing model, slightly different from the one discuss below, and also proves nonparametric identifiability.

these models is helpful for studying more complex ones.

We first prove that without variation in the contracts, the model is not nonparametrically identified. However, the knowledge of one of the primitive functions is sufficient to obtain full identification. Hence, in a regulation context for instance, the model is identified if costs are observed ex post. The identification of first price auctions with risk neutral bidders can also be viewed as a particular case of this result.

To overcome the limitation of knowing one of the primitives, variations in the contracts between the principal and the agents may be used. We study identification when these variations do not stem from changes in the cost function and in the distribution of the agents' type. This amounts to observing an instrumental variable which affects the principal's objective function but not the agent. In the nonlinear pricing model, for instance, one may use any cost shifter of the firm (the principal), since it induces changes in its objective function but is unlikely to affect the consumers directly. Similarly, in the delegation of a production to agents, any demand shifter on the produced good is likely to be a valid instrument. In a first price auction setting, the number of bidders also satisfies this requirement if it is independent of the valuation of the good (Guerrre et al., 2008).

To characterize identification under this condition, we extensively use the first order condition, which defines the optimal choices of the agents, and the link between the observed distribution of the trades and the unobserved distribution of the agents' types. The first equation allows us to define what we call horizontal transforms whereas the second one yields vertical transforms. These transforms are identified in the data and can be combined to identify recursively the functions of interest. Building on this idea, we show that the model is fully identified under a crossing condition. This result is strongly related to the recent result of Guerrre et al. (2008) on the nonparametric identification of risk aversion in first price auctions. We extend here their result to other adverse selection frameworks. We also study identification when the crossing condition does not hold. In these cases, the identification results depend on how much the instrument varies. The model is set identified when the instrument takes two values, but can be point identified once three different values are observed. Finally, when the instrument is continuous, the model is identified without the cost separability condition.

An important feature of our identification procedure is that the cost 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 Sapppington, 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. We also show that our results extend to endogenous variations of the contracts, i.e. when changes in the contracts are partly due to unobserved shocks on the agents' characteristics. Similarly to Guerrre et al. (2008), our previous results still apply if the unobserved shock satisfies a monotonicity condition. Such an assumption is usual in instrumental framework (see e.g. Imbens and Newey, 2009). Finally, we consider two important extensions. First, we show that our recursive method can still be applied when the cost separability condition is replaced by other restrictions. Second, variations of contracts may induce selection effects. We show how to deal with this issue when panel data are available.

Beyond identification, we develop a nonparametric estimation procedure based on our recursive identification method. Our procedure covers both the partial and full identification case. We first recover the horizontal and vertical transforms. We then use them to estimate the primitives on a sequence of points. Finally, these points allow us to build bounds on the structural functions when partial identification occurs, and to point estimate these functions under full identification. This latter case is more difficult to deal with since identification is obtained after an infinite number of iterations, whereas the estimation error increases with the number of iterations. As a consequence, we face a trade-off between bias and variance which is usual in nonparametric estimation. We show that consistency of the estimator can be achieved if the maximal number of iterations allowed in the estimation procedure is finite but tends to infinity with the sample size, at an appropriate rate. Monte Carlo simulations show that the estimators perform well in practice.

Finally, we apply our method to contract data between the French National Institute of Economics and Statistics (Insee) and its interviewers[3] to study incentives in firms in the spirit of Ferrall and Shearer (1999) and Paarsch and Shearer (2000). We have reasons to believe that Insee does not implement the optimal contracts but only optimizes over linear ones. However, our method still applies. Thanks to an exogenous change in the bonus paid to interviewers, we recover bounds on the cost function of the agents and the distribution of their type. Then, using parametric specifications in line with our nonparametric estimates, we compute the cost of using two part tariffs instead of the optimal ones. We find that Insee's loss is about 1%. This result contrasts with Ferrall and Shearer (1999)'s one in

[3] Insee hires interviewers for its household surveys.

which simple contracts were found to lead to a large loss of around 50%. It is however more in line with a recent empirical paper of Miravete (2007) who reports a loss of only 3%. This result also supports the theoretical findings of Wilson (1993, section 6.4), Neeman (2003), Rogerson (2003) and Chu and Sappington (2007).

Finally, we also recover what Insee's surplus would have been under complete information and find that the estimated expected surplus under incomplete information is 83% of the full information surplus. Overall, the cost due to the asymmetry of information is twice the cost associated with the use of a simple bonus system instead of the optimal one.

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 methods and Monte Carlo simulations are presented in section 4 and 5 respectively. The application is displayed in section 6, and section 7 concludes. All proofs are deferred to the appendix.

2 Adverse selection model

Since the seminal work of Myerson (1979, 1981), extensive attention has been devoted to the theoretical properties of adverse selection models. We follow closely here the presentation of Laffont and Martimort (2002) and 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 $U(t, y, \theta) = t - C(y, \theta)$.[4] 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. Our first assumption restricts the functional form of the cost function.

Assumption 1 *(cost separability)* $C(y, \theta) = \theta C(y)$ *where* $C(.)$ *is three times continuously differentiable,* $C'(y) > 0$ *for all* $y > 0$ *and* $C''(.) > 0$.

[4]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. Note also that since the agent's program is determinist, we could suppose him to be risk averse (i.e. $U(t, y, \theta) = u(t - C(y, \theta))$ for a concave increasing function u) without modifying the results of the program. We omit u from the discussion since it is obviously not identified.

Basically, cost separability is a restriction that reduces the dimensionality of the problem. In general, such a condition is necessary to obtain identification results. This assumption is quite common in the theoretical literature (see e. g. Wilson, 1993, or Laffont and Tirole, 1993) as well as in empirical works (see Wolak, 1994, Ferrall and Shearer, 1999, Lavergne and Thomas, 2005). It is also assumed by Perrigne and Vuong (2004) in their nonparametric analysis of the regulation model. We come back to this assumption in subsection 3.3, and show that our results can be extended to other restrictions on the cost function, and even be relaxed in particular settings.

We now analyze separately the agent's program from the firm's one. Indeed, if everyone is usually ready to believe that the agents behave optimally, it is less clear that the econometrician always wants to impose the optimality of the contracts. Hence we differentiate in the paper the results that only rely on the optimal behavior of the agent from those that also require the contracts to be optimal.

2.1 The agent's program

The agent faces a set of contracts of the form $[(y,t(y)); y \in \mathbb{R}^+, t(y) \in \mathbb{R}^+]$. 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 $\underline{U}$. Hence, the agent of type θ chooses the trade $y(\theta)$ satisfying

$$y(\theta) \in \arg\max_{y}[t(y) - \theta C(y)]. \tag{2.1}$$

Moreover, $y(\theta)$ is implemented if and only if the agent participates, ie $\max_y t(y) - \theta C(y) \geq \underline{U}$. In the following, we rely on the first order condition of the agent. For this approach to be valid, the regular conditions below are imposed.

Assumption 2 *(regular conditions 1)* $t(.)$ *is twice differentiable and for all* $\theta \in]\underline{\theta}, \overline{\theta}[$, $t''(.) - \theta C''(.) < 0$, $t'(0) - \theta C'(0) > 0$, $\lim_{y\to+\infty} t'(y) - \theta C'(y) < 0$ *and* $\max_y t(y) - \theta C(y) \geq \underline{U}$.

The main condition is twice differentiability, which rules out kinks in the transfer functions and thus bunching, for which the contracts targeted for different types coincide.[5] The other conditions on $t(.)$ hold for instance if $t(.)$ is concave, $C'(0) = 0$ and $\lim_{y\to\infty} C'(y) = \infty$.

[5]Because bunching leads to rather different results both in theory and in terms of identification, we leave this case for future research.

Under conditions 1-2, every agent participates and the agent's program (2.1) admits a unique solution which is defined by the first order condition

$$t'(y(\theta)) = \theta C'(y(\theta)). \tag{2.2}$$

Moreover, by differentiating this equation, $y'(\theta)$ satisfies

$$[t''(y(\theta)) - \theta C''(y(\theta))]y'(\theta) = C'(y(\theta)).$$

Thus, $y'(\theta) < 0$ and there is indeed no bunching of types.

2.2 The principal's program

Given the agent's program (2.1), the principal chooses the transfer function in order to maximize his objective function. We suppose here the principal to be risk neutral and his objective function to be quasi linear, $W(t, y, \theta) = S(y) - t$.[6] Let $t^*(.)$ denote the optimal contract for the principal. $t^*(.)$ satisfies

$$t^* \in \arg\max_{t(.)} \int [S(y(\theta)) - t(y(\theta))] f_\theta(\theta) d\theta \quad \text{s.t. } y(\theta) \text{ satisfies (2.1).}$$

Without further restriction, $t^*(.)$ does not necessarily satisfy assumption 2. The optimal contract may lead to bunching, for instance. Besides, the first order condition of the principal may neither be necessary nor sufficient to describe the optimal contracts. To avoid these technicalities, we impose the following regularity conditions.

Assumption 3 *(regularity conditions 2) S is twice differentiable with $S'(.) > 0$, $S''(.) < 0$, f_θ is continuously differentiable with $f_\theta(.) > 0$, $\theta \mapsto \theta + F_\theta(\theta)/f_\theta(\theta)$ is strictly increasing and for all $\theta \in]\underline{\theta}, \overline{\theta}[$, $S'(0) - [\theta + F(\theta)/f(\theta)]C'(0) > 0$ and $lim_{y \to +\infty} S'(y) - [\theta + F(\theta)/f(\theta)]C'(y) < 0$.*

These regularity conditions state that the objective function of the principal is increasing and concave, and that the distribution of θ satisfies a technical condition which holds for most single-peaked densities.

Theorem 2.1 *Under assumptions 1 and 3,*

[6] In particular cases such as regulation, it may be relevant to let S also depend on θ. We mention how to adapt our results to this case when necessary.

1. *The trade* $y^*(\theta)$ *corresponding to the optimal transfer is defined by*

$$S'(y^*(\theta)) = \left[\theta + \frac{F_\theta(\theta)}{f_\theta(\theta)}\right] C'(y^*(\theta)). \tag{2.3}$$

2. *The optimal transfer function is defined by* $t^{*\prime}(y^*(\theta)) = \theta C'(y^*(\theta))$ *and the border condition* $t(y^*(\overline{\theta})) - \overline{\theta} C(y^*(\overline{\theta})) = \underline{U}$*. It satisfies assumption 2.*

The theorem is proved for instance in Laffont and Martimort (2002). Integrating (2.2) shows that, compared to the symmetric optimal contract where agents'type is observed by the principal, the firm has to leave a positive information rent $\int_\theta^{\overline{\theta}} C(y^*(\tau))d\tau$ to agents of type θ for them to reveal their types. This information rent increases with the efficiency of the agent and creates inefficiencies in production (the term $F_\theta(\theta)/f_\theta(\theta)C'(y^*(\theta))$ in equation (2.3)). Besides, one can remark that if contracts are optimal, assumption 3 automatically entails assumption 2. Hence, the first-order approach of the agent's program is valid and there is no bunching at equilibrium.

2.3 Examples

There are several classical settings where such an adverse selection model is useful. A first one is price discrimination. In Mussa and Rosen (1978), the principal is a firm that produces a good of quality y at a cost $H(y)$. Agents have heterogenous preferences for quality θ and have a utility $U = \theta y - t$ if they pay t for a good of quality y. The same model can be used to study nonlinear pricing by a monopoly (see e.g. Maskin and Riley, 1984). A second example is financial contracts. In Freixas and Laffont (1990) framework, the principal is a lender who provides the borrower with a loan y. His utility is $S(y) = t - Ry$, where R denotes the risk-free interest rate. Agents are firms with profit $U = \theta f(y) - t$, where $\theta f(y)$ is the production of the firm, y represents the units of capital and θ is a productivity index. A third example is 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 notations,

$$S(y,\theta) - t(y) = (1-\alpha)\left[\int_0^y p(u)du - t(y)\right] + \alpha\left[t(y) - \theta C(y)\right], \tag{2.4}$$

where $p(.)$ denotes the price function.

Auctions can be interpreted as a special case of adverse selection models. The parameter θ (usually denoted v in these models) is the valuation for the good and y corresponds to the bid. In a first price auction with n risk-neutral bidders for example, the utility of the agents takes the form $U(y,\theta) = \theta F_y^{n-1}(y) - yF_y^{n-1}(y)$, where F_y is the cumulative distribution function of y and $F_y^{n-1}(y)$ corresponds to the probability of winning the auction with a bid y. Thus, we recover a separable form of the kind $U(y,\theta) = \theta U(y) - t(y)$ that corresponds to the model.

Finally, this model is also useful for simple insurance settings without moral hazard. The insurance company proposes to agents contracts of the form $(y, d(y))$ where y and $d(y)$ denote respectively the premium and the corresponding deductible. Agents are heterogenous with respect to the probability p of facing an accident. Letting $u(.)$ denote their VnM utility, the expected utility of agent p when choosing y is given by

$$U(y,p) = u(-y) + p\left[u(-y - d(y)) - u(-y)\right]$$

and the first order condition satisfies

$$1 - d'(y) = \frac{p-1}{p}\frac{u'(-y)}{u'(-y-d(y))}.$$

The model fits within the previous framework by letting $t'(y) = 1 - d'(y)$, $C'(y) = \frac{u'(-y)}{u'(-y-d(y))}$ and $\theta = \frac{p-1}{p}$.

3 Nonparametric identification

3.1 The setting

We now turn to the empirical content of the model. We suppose that the econometrician observes the trades at equilibrium $y_{X,Z}(\theta)$ for an infinite sample of agents and for several menus of contracts indexed by covariates X and instruments Z whose role is clarified below. We also suppose that the corresponding transfers $t_{X,Z}(y_{X,Z}(\theta))$ are observable, for each menu of contracts.[7] The trades and transfers enable one to identify the cumulative distribution function of $y_{X,Z}(\theta)$, $F_{y_{X,Z}}(.)$ and the transfer function $t_{X,Z}(.)$ on the support $\mathcal{Y}_{X,Z}$ of $y_{X,Z}(\theta)$. The question is whether the cost functions $C_{X,Z}(.)$, the distributions of types $F_{\theta,X,Z}(.)$ and the principal's objective function $S_{X,Z}(.)$ can be recovered from these functions and the model.

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

In general, variation in the menus of contracts are due to changes in the cost function, the distribution of types or the principal objective function. Without exclusion restriction, such variations do not have any effect on identification since only the data for $X = x$ and $Z = z$ can be used to identify $C_{x,z}$, $F_{\theta,x,z}$ and $S_{x,z}$. We impose here the exclusion restriction $(C_{X,Z}, F_{\theta,X,Z}) = (C_X, F_{\theta,X})$. In other terms, Z does not affect neither the cost function nor the distribution of types. Variation in the contracts only stems from a change in the principal's function through a change in Z.

Except in subsection 3.2.5, we suppose that Z only takes a finite number of values. Because the value of the instrument does not play any role, we can suppose without loss of generality that $Z \in \{1, ..., K\}$. Finally we suppress X in our notations for the ease of exposition. Agents are thus supposed to be homogenous except for their unknown types; if they differ by observed characteristics, our results below must be understood to be conditional on these characteristics.[8] With these notations, the exclusion restriction takes the following form:

Assumption 4 *(exclusion restriction)* $C_1 = ... = C_K = C$ *and* $F_{\theta,1} = ... = F_{\theta,K} = F_\theta$.

There are several situations where assumption 4 is likely to hold. In the monopoly price discrimination model, the price of an input may vary, inducing a change in the cost function of the monopoly and thus in S. However, this variation does not affect the utility function of the consumer. As usual in this literature, any cost shifter Z may play the role of the instrument. Similarly, in the delegation of a production to agents, exogenous variations of the market value of the product affects the principal's objective function but not the agents' one. In such cases, any demand shifter is a valid instrument Z. An example of this kind is developed in section four. In the regulation context, one may use for example changes in the government color, which induce variation in the parameter α (Gagnepain and Ivaldi, 2007) but not in (F_θ, C). Finally, in an auction context, one may rely on changes in the number of bidders, following Guerrre et al. (2008).

Our framework can also be applied even if the changes in the menus are due to modifications of C and F_θ. Indeed, suppose that these modifications appear continuously, while the principal only modifies his menu of contracts from time to time, because of menu costs. This situation typically arises in nonlinear pricing or price discrimination. Then trades and transfers just before and after the menu change satisfy assumption 4. This idea is close to the one of regression discontinuity (see e.g. Hahn et al., 2001). In this case,

[8] The case where they differ by unobserved characteristics is postponed until subsection 3.3.1.

the menu of contracts just before the change is inoptimal in the sense that it does not correspond defined in Theorem **??**. However, as will become clear below, this does not preclude identification of C and F_θ.

Other examples involving inoptimal variations of contracts are experiments, in which different menus of contracts are proposed to people in a random way. 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 in the duration of the loan.

3.2 Identification results

Before turning to our results, note that a normalization is necessary since we can replace $(\theta, C(.))$ by $(\alpha\theta, C(.)/\alpha)$ and leave the model unchanged. Hence, for a given $y_0 \in \mathcal{Y}_1$, we can choose any $\theta_0 > 0$ such that $\theta_1(y_0) = \theta_0$.[9,10] Identification is based on three equations. First, we have shown that under assumption 2, $y_k(.)$ admits an inverse $\theta_k(.)$. This function satisfies, for all $y \in \mathcal{Y}_k$,

$$1 - F_{y_k}(y) = \mathbb{P}(y_k(\theta) \geq y) = \mathbb{P}(\theta \leq \theta_k(y)) = F_\theta(\theta_k(y)). \tag{3.1}$$

The first equality stems from the fact that the distribution of y is atomless and the second from $\theta_k(.)$ being strictly decreasing.

We also rely on the agent's program for identification. Following Guerre et al. (2000), we use the first-order condition of the agent, which defines the unique solution of the program under assumption 2. Taking equation (2.2) at $\theta_k(y)$, this condition satisfies

$$t_k'(y) = \theta_k(y) C'(y). \tag{3.2}$$

Finally, if one is willing to assume optimality of contracts, the first order condition of the principal (2.3) can be used. It is convenient to rewrite this condition in terms of the

[9]We recommend to make the normalization for a y_0 in the interior of $\mathcal{Y}_1 = [\underline{y}_1, \overline{y}_1]$. Indeed, if $\overline{\theta} = +\infty$, then normalizing $\theta_1(\underline{y}_1)$ to a finite value leads to inconsistent results. The same holds if $\underline{\theta} = 0$ and one chooses to normalize $\theta_1(\overline{y}_1)$ to a positive value. Other recommendations on the choice of y_0 are given subsequently.

[10]Once a normalization has been done on $\theta_1(.)$, no other normalization on the $\theta_k(.)$, $k \geq 2$, is needed. This is because the normalization on $\theta_1(y_0)$ induces a normalization on $C'(.)$ and F_θ. This normalization then applies to all other menus of contracts.

observables. Using (3.1) and (3.2), we obtain

$$S'_k(y) = \left[1 - \frac{1 - F_{y_k}(y)}{f_{y_k}(y)} \frac{\theta'_k(y)}{\theta_k(y)}\right] t'_k(y), \qquad (3.3)$$

where $f_{y_k}(.)$ denotes the density of $y_k(\theta)$.

In the following, we first provide identification results without variation in the menu of contracts ($K = 1$). We then introduce the horizontal and vertical transforms which are at the core of our identification strategy when $K \geq 2$. Finally, we present our results for $K = 2$, $K \geq 2$ and when the instrument is continuous.

3.2.1 Identification result for $K = 1$

Without variation in the menus of contracts, we have in hand the three equations (3.1), (3.2) and (3.3) whereas we seek to recover the four functions $C'(.)$, $F_\theta(.)$, $S'_1(.)$ and $\theta_1(.)$. Thus we can expect the model to be identified only up to one of these functions. Theorem 3.1 formalizes this intuition.

Theorem 3.1 *Suppose that $K = 1$, the menu of contracts is optimal and assumptions 1 and 3 hold. Then (C', F_θ, S'_1) are not identified jointly. On the other hand, if one of these three functions is known, the other two can be identified.*

If the first part of the theorem is not surprising, the second part has some interesting consequences. It can be applied to the regulation example, even though S depends on θ in this case. Without imposing any form on the principal's objective function, the model is identified if costs are observable ex post, as in Wolak (1994) and Perrigne and Vuong (2004). Hence, we can also test the optimality of contracts under the supplementary condition that the principal's objective function satisfies (2.4).[11] In the auction framework, Theorem 3.1 simply states that the first price auctions model with risk neutral bidders is identified as the probability of winning $F_y^{n-1}(y)$ that appears in the agent condition is known (Guerre et al., 2000). Another example is the price discrimination model. If, following Crawford and Shum (2007), we specify the utility of the agent by $U = \theta q - t$, the rest of the model is nonparametrically identified.[12] Finally, we show in subsection 3.3.3 that this theorem is useful to solve selection effects that may appear when $K \geq 2$.

[11] This result differs slightly from the one of Perrigne and Vuong (2004). This is because they consider the regulation model of Laffont and Tirole (1986) where the regulated firm makes a costly and unobserved effort to reduce its cost. Thus, they have to deal both with adverse selection and moral hazard.

[12] In addition to this specification, Crawford and Shum (2007) need to specify the principal's objective function as they do not observe the quality q.

Nevertheless, Theorem 3.1 can be sometimes restrictive since one of the functions must be known. Moreover, one of the main issue in the empirical literature on contract theory is whether the observed contracts are optimal (Chiappori and Salanié, 2002). Answering this question requires estimating the principal's objective function and comparing it with the theoretical one. Usually, this cannot be achieved when $K = 1$. As we will see, observing different menus of contracts under our exclusion restriction overcomes these limitations, by providing identification even when none of the functions is known.

3.2.2 The horizontal and vertical transforms

We now introduce two types of transforms that are at the basis of our identification method when $K \geq 2$. First, equation (3.1) implies that for all $\theta \in \Theta$,

$$F_{y_i}(y_i(\theta)) = F_{y_j}(y_j(\theta)).$$

Hence, letting $H_{ij}(y) = F_{y_j}^{-1}\left[F_{y_i}(y)\right]$ denote the quantile-quantile transformation between the distribution of $y_j(\theta)$ and $y_i(\theta)$, we get

$$y_j(\theta) = H_{ij}(y_i(\theta)). \tag{3.4}$$

Because $H_{ij}(.)$ is identified on $\mathcal{Y}_i$, the knowledge of $y_i(\theta)$ implies the knowledge of $y_j(\theta)$. The horizontal transforms on figure 1 are thus identified and we can recover point (1) for instance if we know point (0). Another consequence is that it suffices to identify $y_1(.)$ (or equivalently $\theta_1(.)$) to recover the other functions $y_k(.)_{2\leq k\leq K}$ (or $\theta_k(.)_{2\leq k\leq K}$).

Second, suppose that the intersection of the supports $\mathcal{Y}_i$ and $\mathcal{Y}_j$ is not empty and let $y \in \mathcal{Y}_i \cap \mathcal{Y}_j$. The first order condition implies that

$$\frac{t_i'(y)}{\theta_i(y)} = \frac{t_j'(y)}{\theta_j(y)}.$$

If we define the vertical transform $V_{ij}(.,.)$ by $V_{ij}(\theta, y) = t_j'(y) \times \theta / t_i'(y)$, we get

$$\theta_j(y) = V_{ij}(\theta_i(y), y). \tag{3.5}$$

Because $V_{ij}(.,.)$ is identified on $\mathbb{R} \times \mathcal{Y}_i \cap \mathcal{Y}_j$, the knowledge of $\theta_i(y)$ implies the knowledge of $\theta_j(y)$. Starting from point (1) on figure 1, for instance, we can identify point (2).

To conclude, starting from $(y_0, \theta_1(y_0))$, we can identify $(y_1, \theta_1(y_1))$ where $y_1 = H_{12}(y_0)$ and $\theta_1(y_1) = V_{21}(\theta_1(y_0), y_1)$. By induction, we identify all the black points in figure 1.

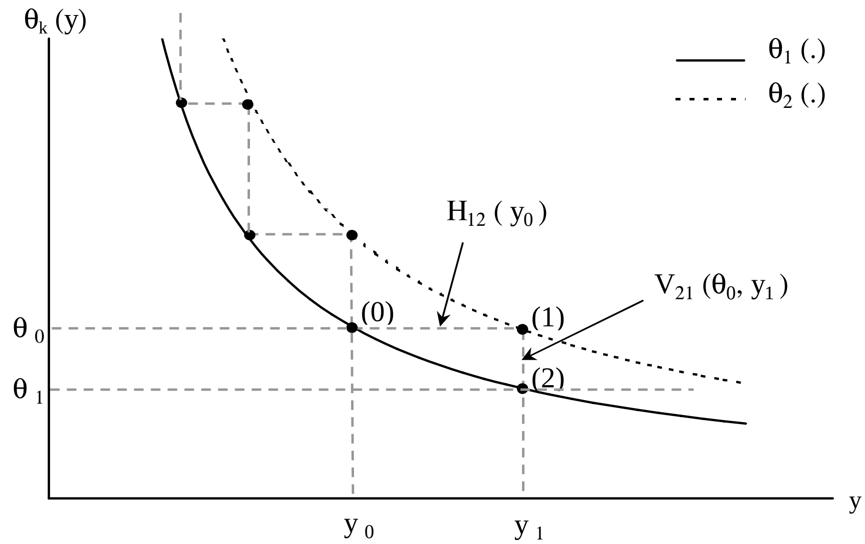

Figure 1: The horizontal and vertical transforms.

3.2.3 Identification results for $K = 2$.

Figure 1 corresponds to a situation where $\theta_1(y) < \theta_2(y)$ for all $y \in \mathcal{Y}_1 \cap \mathcal{Y}_2$. This does not hold in general, since $\theta_1(.)$ and $\theta_2(.)$ may cross. Actually, because $\theta_1(y) < \theta_2(y)$ is equivalent to $t'_1(y) < t'_2(y)$ by equation (3.2), the two cases can be distinguished in the data. We consider them separately as they lead to different results in terms of identification.

Figure 1 suggests that without crossing, $\theta_1(.)$ can be identified on a sequence. Formally, we identify $\theta_1(.)$ on the increasing sequence $(y_n)_{n \in I_1}$ defined on a subset I_1 of $\mathbb{Z}$ as follows. First, let $0 \in I_1$. Then $n \in I_1$ if $n - 1 \geq 0$ belongs to I_1 and $H_{12}(y_{n-1}) \in \mathcal{Y}_1$. In this case, $y_n = H_{12}(y_{n-1})$. Otherwise, $n \notin I_1$. Similarly, if $n + 1 \leq 0$ belongs to I_1 and $y_{n+1} \in \mathcal{Y}_2$, $n \in I_1$ and $y_n = H_{21}(y_{n+1})$. Otherwise once more, $n \notin I_1$.

By monotonicity of $\theta_1(.)$, $\theta_1(y) \geq \underline{\theta}_1(y) = \sup_{n \in I_1 : y_n \geq y} \theta_1(y_n)$ and $\theta_1(y) \leq \overline{\theta}_1(y) = \inf_{n \in I_1 : y_n \leq y} \theta_1(y_n)$. Similarly, $y_1(.)$ is bounded by $\underline{y}_1(\theta) = \sup_{n \in I_1 : \theta_1(y_n) \geq \theta} y_n$ and $\overline{y}_1(\theta) = \inf_{n \in I_1 : \theta_1(y_n) \leq \theta} y_n$.[13] Then, using equations (3.1), (3.2), (3.3) and monotonicity conditions, we are able to identify bounds on $F_\theta(.)$, $C'(.)$ and $S'_k(.)$ (with $k = 1, 2$) respectively, as Theorem 3.2 shows.

Theorem 3.2 *Suppose that $K = 2$, assumptions 1, 2 and 4 hold and $t'_1(.) < t'_2(.)$. Then:*

[13] When the set $\{n \in I_1 : y_n \geq y\}$ (resp. $\{n \in I_1 : \theta_1(y_n) \leq \theta\}$) is empty, we let $\underline{\theta}_1(y) = 0$ (resp. $\underline{y}_1(\theta) = \underline{y}_1$). The same holds with $\overline{\theta}_1(y)$ and $\overline{y}_1(\theta)$, replacing 0 and $\underline{y}_1$ by respectively $+\infty$ and $\overline{y}_1$.

1. *$C'(.)$ and $F_\theta(.)$ are point identified on respectively $(y_n)_{n\in I_1}$ and $(\theta_1(y_n))_{n\in I_1}$ and partially identified elsewhere, with:*

$$
\begin{aligned}
1-F_{y_1}(\overline{y}_1(\theta)) = \underline{F_\theta}(\theta) &\le F_\theta(\theta) \le \overline{F_\theta}(\theta) = 1-F_{y_1}(\underline{y}_1(\theta))), \qquad (3.6)\\
\frac{t'(y)}{\overline{\theta}_1(y)} = \underline{C'}(y) &\le C'(y) \le \overline{C'}(y) = \frac{t'(y)}{\underline{\theta}_1(y)}. \qquad (3.7)
\end{aligned}
$$

2. *If one replaces assumption 2 by assumption 3 and if contracts are optimal, $S_k(.)$ satisfies, for all $y > y'$*

$$
\begin{aligned}
S_k(y)-S_k(y') &\ge t_k(y)-t_k(y') + \min_{u\in[y',y]} q_k(u)\left[\ln\underline{\theta}_1(y') - \ln\overline{\theta}_1(y)\right],\\
S_k(y)-S_k(y') &\le t_k(y)-t_k(y') + \max_{u\in[y',y]} q_k(u)\left[\ln\overline{\theta}_1(y') - \ln\underline{\theta}_1(y)\right]
\end{aligned}
$$

where $q_k(y) = t'_k(y)(1-F_{y_k}(y))/f_{y_k}(y)$.

Theorem 3.2 provides bounds on $C'(.)$ and $F_\theta(.)$ when $t'_1(.)$ and $t'_2(.)$ do not cross. Actually, the bounds on $C'(.)$ and $F_\theta(.)$ coincide on $(y_n)_{n\in I_1}$ and $(\theta_1(y_n))_{n\in I_1}$ respectively. The sequences $(C'(y_n))_{n\in I_1}$ and $(F_\theta(\theta_1(y_n)))_{n\in I_1}$ may serve as a guidance for choosing appropriate parametric restrictions. Beyond accuracy reasons, such restrictions are necessary to recover parameters of interests when contracts are inoptimal (see our application in section 6). When contracts are optimal, $S_k(.)$ $(k=1,2)$ is also partially identified, up to a constant.[14] However, the lower and upper bounds do not coincide on any value of y in general. Intuitively, fewer information can be obtained on this function than on (C', F_θ) since it does not only depend on $\theta_k(.)$ but also on $\theta'_k(.)$, which is nowhere point identified.[15]

We now turn to the case where the functions $t'_1(.)$ and $t'_2(.)$ cross.[16] In this case, the model can be fully recovered thanks to the intersection points. The proof is quite different from previously and can be explained as follows. By the normalization, the value $\theta_1(y_0)$ of an intersection point y_0 of $t'_1(.)$ and $t'_2(.)$ can always be fixed to any θ_0. For any y_α and θ^0, define the sequence $(\theta^n)_{n\in\mathbb{N}}$ as in figure 2. We show that $(\theta^n)_{n\in\mathbb{N}}$ always converges, but reaches θ_0 only if $\theta^0 = \theta_1(y_\alpha)$. This enables to recover $\theta_1(y_\alpha)$, since θ_0 is known. Because

[14] We obtain bounds on differences on $S_k(.)$ rather than on the function itself since the first order condition only brings information on $S'_k(.)$. Besides, note that such bounds cannot be obtained when S_k also depends on θ.

[15] In general, the bounds on $C'(.)$, F_θ and $S'_k(.)$ are not sharp because they can violate the regularity conditions imposed in assumption 1, 2 or 3. We show in the supplementary material how a full use of these conditions may yield better bounds.

[16] For the sake of simplicity, we consider the case where transfer functions cross at a finite set of points.

y_α was arbitrary, this proves that $\theta_1(.)$ is fully identified. Thus, all the functions of interest can be recovered.[17]

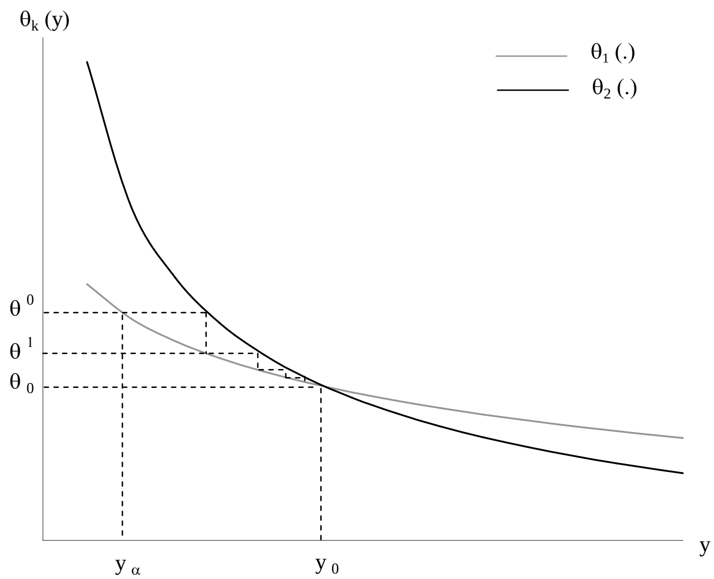

Figure 2: Identification when $t_1'(.)$ and $t_2'(.)$ cross.

Theorem 3.3 *Suppose that $K = 2$, assumptions 1, 2 and 4 hold and $t_1'(y) = t_2'(y)$ for a finite set of points $y_0 < ... < y_M$ in the interior $\overset{\circ}{\mathcal{Y}}_1$ of $\mathcal{Y}_1$. Then $C'(.)$ and $F_\theta(.)$ are identified on $\mathcal{Y}_1 \cup \mathcal{Y}_2$ and Θ respectively. If one replaces assumption 2 by assumption 3 and if contracts are optimal, $S_1'(.)$ and $S_2'(.)$ are also identified, on $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively.*

Theorem 3.3 is reminiscent of the result of Guerrre et al. (2008) in the context of first-price auctions with risk averse bidders. They use exogenous variations (namely, variation in the number of bidders) to obtain identification of the model at the limit, using a converging sequence (see their proposition 1). We provide further details on the link between Guerrre et al. (2008)'s result and Theorem 3.3 in subsection 3.3.2.

3.2.4 Identification results for $K \geq 3$.

If two of the K functions $t_k'(.)$ cross, we can apply the previous result and the model is identified. Thus, in this subsection, we study the noncrossing case with two or more

[17]When S_k depends on θ, one can only identify $\frac{\partial S_k}{\partial y}(y, \theta_k(y))$. This does not able to recover S in general. However, this still allows to test the optimality of contracts by comparing the theoretical $\frac{\partial S_k}{\partial y}(y, \theta_k(y))$ with the one obtained with the data. The same remark applies to Theorem 3.5 and Theorem 3.6 below.

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 $\theta_1(.)$ is identified is larger. Figure 3 gives an example where starting from (y_0, θ_0) on the curve $\theta_1(.)$, we can identify $\theta_1(.)$ on y_1 as previously but also between y_0 and y_1 (on y_2 for instance). Theorem 3.4 defines the precise set where the function $\theta_1(.)$ is point identified.

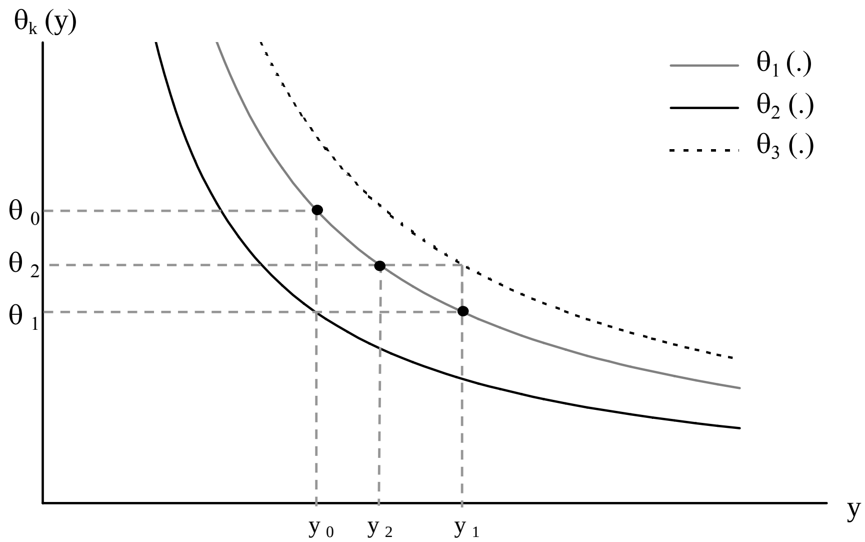

Figure 3: Identification with $K = 3$ different transfer functions.

Theorem 3.4 *Suppose that $K \geq 3$, assumptions 1, 2 and 4 hold and $t_1'(.) < ... < t_K'(.)$. $\theta_1(.)$ is identified on $\overline{\mathbb{Y}} \cap \mathcal{Y}_1$ where $\overline{\mathbb{Y}}$ is the closure of the set $\mathbb{Y}$ defined by:*

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

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 $F_\theta(.)$. 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 5 *(separability in the transfer functions) For all y,*

$$t_k'(y) = \delta_k m(y),$$

where we let (without loss of generality) $\delta_2 < \delta_1 < \delta_3$.

Assumption 6 *(non periodicity): there exists* $1 \leq i < j < k \leq K$ *such that*

$$\frac{\ln(\delta_i/\delta_j)}{\ln(\delta_k/\delta_j)} \notin \mathbb{Q}. \tag{3.8}$$

We let (without loss of generality) $i = 1$, $j = 2$ *and* $k = 3$.

Assumption 7 *(large support 1):* $H_{12}(\mathcal{Y}_2 \cap \mathcal{Y}_3) \cap H_{13}(\mathcal{Y}_2 \cap \mathcal{Y}_3) \cap \mathcal{Y}_1 \neq \emptyset$.

Assumption 6 is a technical condition which ensures that the identifying sequences are not periodic. Because almost every real is 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 $\theta_1(.)$ is identified. This assumption can be easily checked in the data. The more limiting assumption is the separability condition, which is, up to a reindexation, a restriction of $t'_1(.) < ... < t'_K(.)$. Nevertheless, it includes the common case of constant marginal transfers. Moreover, it is directly testable in the data.[18] Theorem 3.5 shows that the model is fully nonparametrically identified under these assumptions.

Theorem 3.5 *Suppose that* $K \geq 3$ *and assumptions 1, 2 and 4-7 hold. Then* $C'(.)$ *and* $F_\theta(.)$ *are identified on* $\cup_{k=1}^K \mathcal{Y}_k$ *and* Θ *respectively. Moreover, if one replaces assumption 2 by assumption 3 and if contracts are optimal,* $S'_k(.)$ $(k = 1, .., K)$ *is also identified on* $\mathcal{Y}_k$.

Theorem 3.5 implies that when $K \geq 4$, the cost separability assumption can be tested (the other assumptions being maintained). Indeed, we can recover F_θ and C' by using different subsets of menus. If the different corresponding functions do not coincide, this assumption is rejected.

3.2.5 Identification result with a continuous instrument

In some settings, the instrument may be continuous. In the price discrimination example, the price of an input of the monopoly may take any value in an interval, implying that the value function of the principal changes continuously. As mentioned above, the model with cost separability is now overidentified. Theorem 3.6 shows that actually, full identification can be obtained without this assumption. In other terms, no restriction is needed on C, defined as a function of two variables $C(y, \theta)$. In this case, we can always normalize θ to be uniform.

[18] When contracts are optimal, this condition is satisfied if there exists a function $Q(.)$ such that for all $k = 1...K$, $S_k(y) = G(\delta_k Q(y))\, C'(y)$, with $G(\theta) = \theta + F_\theta(\theta)/f_\theta(\theta)$. However, as mentioned before, the optimality of contracts is not required to recover $C'(.)$ and $F_\theta(.)$.

Theorem 3.6 *Suppose that z takes values in $\mathcal{Z} = [\underline{z}, \overline{z}]$, θ is uniformly distributed on $[0,\ 1]$, $(y, z) \mapsto t(y, z)$ is twice differentiable and $\frac{\partial^2 t}{\partial y \partial z}(y, z) > 0$.*[19] *Then $\frac{\partial C}{\partial y}(.,.)$ is identified on $\{(y, \theta) : \theta \in \Theta,\ \exists z \in \mathcal{Z} : \theta(y, z) = \theta\}$. Moreover, if contracts are optimal, $\frac{\partial S}{\partial y}(., z)$ is identified on $\mathcal{Y}_z$.*

3.3 Discussion and extensions

3.3.1 Endogenous changes in the menus of contracts

Up to now, we have discussed cases where menus of contracts change exogenously, according to observable instruments Z.[20] In practice, it may happen that a random term ε observed by the principal but not by the econometrician affects the cost function of the agents or their distribution of types. In this case, contracts change endogenously and the transfer function depends not only on Z but also on ε. Similarly to Guerre, Perrigne and Vuong (2008) when considering endogenous participation to auctions, our method can still be applied, under the exclusion restriction that $C_{Z,\varepsilon} = C_\varepsilon$ and $F_{\theta,\varepsilon,Z} = F_{\theta,\varepsilon}$. Suppose that

$$t'_{Z,\varepsilon}(y) = \psi(y, Z, \varepsilon),$$

where ψ is strictly monotonic in ε and $\varepsilon \perp\!\!\!\perp Z$. This latter condition is usual in instrumental variable models (see e.g. Imbens and Newey, 2009). It is also natural in our framework for which Z is a determinant of the principal's objective function whereas ε only affects the agent's side.

Under these two conditions, ε is identified up to a strictly increasing function. Indeed, if we suppose that ε is uniformly distributed on $[0, 1]$, then $\Pr(t'_{Z,\varepsilon}(y) \leq t | Z = z) = \Pr(\varepsilon \leq \psi^{-1}(y, Z, t)) = \psi^{-1}(y, Z, t)$, where $\psi^{-1}(y, Z, .)$ is the inverse of $\psi(y, Z, .)$. Hence, ε can be seen as a supplementary covariate and our method applies conditionally on ε. This idea is close to the one of control variables in identification of nonparametric models with endogenous variables (see e.g. Imbens and Newey, 2009). In this framework, the causal effect of the endogenous variable (here the marginal transfer function) on a dependent variable is identified by adding a control variable (here ε) resulting from a first step regression on instruments (here Z).

[19] We also assume that the first-order conditions of the agent and the principal (if contracts are optimal) are necessary and sufficient for optimality. For this to be satisfied, assumptions 2 and 3 could be modified in order to take into account the non-separability of the cost function.

[20] As previously, we omit covariates X for the ease of exposition.

As an illustration, consider the case where ε is an heterogeneity term on the cost function only, and $\varepsilon \mapsto C'_\varepsilon$ is increasing. Then, by, the principal's first order condition,

$$\theta_{Z,\varepsilon}(y) = G^{-1}\left(\frac{S'_Z(y)}{C'_\varepsilon(y)}\right)$$

where $G(\theta) = \theta + F_\theta/f_\theta(\theta)$. Thus, the optimal contract $t^*_{Z,\varepsilon}$ satisfies

$$t^{*}_{Z,\varepsilon}{}'(y) = G^{-1}\left(\frac{S'_Z(y)}{C'_\varepsilon(y)}\right) C'_\varepsilon(y).$$

If $\theta \mapsto F_\theta/\theta f_\theta(\theta)$ is increasing,[21] we can show that $t^{*}_{Z,\varepsilon}{}'$ is strictly increasing in ε, and our previous result applies.

3.3.2 The cost separability assumption

Our results with a discrete instrument are based on the cost separability assumption $C(\theta, y) = \theta C(y)$. In some settings however, other restrictions may be more natural. One example is the delegation of a task to an agent, as in, e.g., Ferrall and Shearer (1999) or Paarsch and Shearer (2000). Suppose that his production depends on an heterogeneity term θ that he observes ex ante, and on his effort e, so that $y = g(\theta, e)$. θ may represent the agent's productivity or the difficulty of the task itself. g is supposed to be increasing in e and known (or specified) by the econometrician. The cost $C(e)$ only depends on the effort e. Because θ is observed ex ante by the agent, this model is not a moral hazard model but a truly adverse selection one.[22] We can reformulate it in our framework by replacing e by $g_2^{-1}(\theta, y)$, where $g_2^{-1}(\theta, .)$ denotes the inverse function of $g(\theta, .)$. In this case, the cost function satisfies the restriction $C(y, \theta) = C(g_2^{-1}(\theta, y))$.

Identification of this model can be obtained along the same lines as above. Letting $e_k(\theta)$ denote the effort chosen by agent θ when facing menu k, we have

$$e_j(\theta) = g_2^{-1}\left[\theta, F_{y_j}^{-1} \circ F_{y_i}\left(g(\theta, e_i(\theta))\right)\right].$$

This defines the horizontal transform here. Besides, by the agents' first order condition,

$$t'_i\left[g\left(\theta_i(e), e\right)\right] \frac{\partial g}{\partial e}(\theta_i(e), e) = t'_j\left[g\left(\theta_j(e), e\right)\right] \frac{\partial g}{\partial e}(\theta_j(e), e),$$

where $\theta_k(.)$ denotes the inverse of $e_k(.)$. Solving this equation in $\theta_j(e)$ defines the vertical transform.[23] Hence, we get similar results as previously. In particular, Theorem 3.5 also

[21] This condition is satisfied for instance by all Fréchet, Weibull and Pareto distributions.

[22] It is sometimes referred to as a "false moral hazard" model (see e.g. Laffont and Martimort, 2002).

[23] Of course, assumptions on the primitives are needed to ensure the unicity of the solution, as well as the validity of the first-order approach.

holds if marginal transfers are constant and the production function is separable, $g(\theta, e) = h_1(\theta)h_2(e)$, as with Cobb-Douglas functions.

Another example of a restriction different from the cost separability condition arises in the first price auction model with risk averse bidders. In this model, the expected utility of the bidders satisfies $U(y, \theta) = F_y^{n-1}(y)u(\theta - y)$ if the player with valuation θ bids y, $u(.)$ denoting the vNM utility of the player. Here the restriction is that the function $U(y, \theta)/F_y^{n-1}(y)$ only depends on $\theta - y$. The first order condition of the agent satisfies (see Guerrre et al., 2008)

$$\theta_n(y) = y + \left(\frac{u}{u'}\right)^{-1}(L(y)),$$

where $L(y) = \frac{G(y)}{(n-1)g(y)}$. When $K = 1$ (i.e., without variation in the number of bidders), the model is not identified because both F_θ and u/u' have to be recovered and Theorem 3.1 applies. With exogenous variation in the number of bidders, our results should be adapted. The horizontal transform is defined as usual but the vertical transform is replaced here by a "diagonal" transform $D(.)$ defined by

$$\theta_{n_2}(D(y)) = \theta_{n_1}(y) + D(y) - y$$

where n_1 and n_2 are two different number of bidders. Guerrre et al. (2008) prove that the functions $\theta_{n_i}(.)$ cross and use a recursive reasoning similar to the one of Theorem 3.3 to prove that the first price auction model is identified as soon as one exogenous change (i.e., two different number of bidders) is observed.

3.3.3 Selection effects

We have supposed until now that variation in the transfer functions does 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 workers. More generally, these effects may arise in competitive environments where agents can choose between several menus of contracts proposed by different principals. In this case indeed, a change in one principal's menu may induce some agents with particular θ to choose the new menu of contracts. Such effects are not taken into account in our model where all types of agent participate in all menus of contracts.[24] Hence, our analysis is not valid in general when selection occurs.

[24]Selection effects could be modellized by letting the participation constraint of agent θ depends on θ.

However, selection effects can be detected if panel data are available. It suffices indeed to compare the distributions of the stayers' and entrants' type, as in Lazear (2000). Moreover, our method still applies even in the presence of selection effects, provided that the distribution $\widetilde{F}_\theta$ of the stayers (i.e., those who participate in all menus of contracts) remains the same for the different menus. Indeed, the exclusion restrictions remains valid on the population of stayers. As a consequence, $\widetilde{F}_\theta$ can be recovered as well as the cost function $C(.)$. And, once $C(.)$ has been identified, we can use data on the movers to recover their distribution of types, by Theorem 3.1. As a consequence, the distribution of the types of the whole population of agents (i.e., movers and stayers) is identified, and we can recover the principal's objective function under optimality of contracts.

4 Estimation

In this section, we define and study estimators of the primitive functions when data on K menus of contracts are available. More precisely, for menu $j \in \{1, ..., K\}$, we suppose to observe the transfer function $t_j(.)$ and a sample $(y_{ij} \equiv y_j(\theta_{ij}))_{i \in \{1,...,N_j\}}$ of N_j trades. We study the behavior of our estimators when $N = \min(N_1, ..., N_K) \to \infty$ and under the following standard assumption of independent sampling.

Assumption 8 *(independent sampling) For all $j = 1, ..., K$, $(\theta_{1j}, ..., \theta_{N_j j})$ are independently drawn from F_θ.*

We also restrict to separable transfer functions, i.e. functions satisfying assumption 5. When $K = 2$, the analysis would remain valid for any continuously differentiable transfer functions, as soon as $t_1'(.) < t_2'(.)$.[25] However, we require it when $K \geq 3$ to apply the full identification result of Theorem 3.5.

Our estimation method follows closely the identification proof. We first define estimators of the sequences of points in $\mathcal{Y}_1$ and Θ which are identified by induction. The set or point estimators of the primitive functions are then built upon these sequences, in a slightly different way depending on whether $K = 2$ or $K \geq 3$.

[25] The case where transfer functions cross would not fit in our estimation procedure. The main reason is that the identification relies on a result at the infinity, which is very different from the identification proof in the noncrossing case.

4.1 Estimation with two menus of contracts

When $K = 2$, we identify the function $\theta_1(.)$ on the increasing sequence $(y_n)_{n\in I_1}$ defined in subsection 3.2. To estimate this sequence, let, for $j \in \{1,2\}$, $\widehat{F}_{y_j}$ (resp. $\widehat{F}_{y_j}^{-1}$) denote the empirical distribution function (resp. empirical quantile function) of the $(y_{ij})_{i\in\{1,\dots,N_j\}}$. Our estimator of H_{ij} is

$$\widehat{H}_{ij}(x) = \widehat{F}_{y_j}^{-1} \circ \widehat{F}_{y_i}(x).$$

Then we estimate I_1 and $(y_n)_{n\in I_1}$ by induction. First, we let $0 \in \widehat{I}_1$ and $\widehat{y}_0 = y_0$. Second, if $n-1 \geq 0$, $n-1 \in \widehat{I}_1$ and $\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}}_1 = [\min_i y_{i1}, \max_i y_{i1}]$, then $n \in \widehat{I}_1$ and $\widehat{y}_n = \widehat{H}_{12}(\widehat{y}_{n-1})$. Otherwise $n \notin \widehat{I}_1$. Similarly, if $n+1 \leq 0$, $n+1 \in \widehat{I}_1$, $\widehat{y}_{n+1} \in \widehat{\mathcal{Y}}_2 = [\min_i y_{i2}, \max_i y_{i2}]$, then $n \in \widehat{I}_1$ and $\widehat{y}_n = \widehat{H}_{21}(\widehat{y}_{n+1})$.[26]

Because $V_{21}(\theta, y) = \delta_1/\delta_2 \times \theta$ when the transfer functions are separable, $\theta_n = \theta_1(y_n)$ satisfies $\theta_n = (\delta_1/\delta_2)^n\theta_0$ for all $n \in I_1$. Thus, we simply estimate it by $\widehat{\theta}_n = (\delta_1/\delta_2)^n\theta_0$ for all $n \in \widehat{I}_1$. The estimator is perfect for all $n \in \widehat{I}_1 \cap I_1$. $(\widehat{\theta}_n)_{n\in\widehat{I}_1}$ differs in general from $(\theta_n)_{n\in\widehat{I}_1}$ only because in general, $\widehat{I}_1 \neq I_1$. Then, following their definition, we estimate $\underline{\theta}_1(.)$ and $\overline{\theta}_1(.)$ by respectively $\widehat{\underline{\theta}}_1(y) = \sup_{n\in\widehat{I}_1:\widehat{y}_n\geq y} \widehat{\theta}_n$ and $\widehat{\overline{\theta}}_1(y) = \inf_{n\in\widehat{I}_1:\widehat{y}_n\leq y} \widehat{\theta}_n$. Similarly we let $\widehat{\underline{y}}_1(\theta) = \sup_{n\in\widehat{I}_1:\widehat{\theta}_n\geq\theta} \widehat{y}_n$ and $\overline{y}_1(\theta) = \inf_{n:\widehat{\theta}_n\leq\theta} \widehat{y}_n$.

Let us now turn to the estimation of F_θ, C', S_1 and S_2. As shown previously, F_θ and C' are point identified respectively on $(\theta_n)_{n\in\mathbb{Z}}$ and $(y_n)_{n\in\mathbb{Z}}$. We estimate them by

$$\begin{aligned} \widehat{F_\theta(\theta_n)} &= 1 - \widehat{F}_{y_1}(\widehat{y}_n), \\ \widehat{C'(y_n)} &= \frac{t_1'(\widehat{y}_n)}{\widehat{\theta}_n}. \end{aligned}$$

Elsewhere, bounds can be obtained on these functions. Following (3.7) and (3.6), we let $\widehat{\underline{C'}}(y) = t_1'(y)/\widehat{\overline{\theta}}_1(y)$, $\widehat{\overline{C'}}(y) = t_1'(y)/\widehat{\underline{\theta}}_1(y)$, $\widehat{\underline{F}}_\theta(\theta) = 1 - \widehat{F}_{y_1}(\widehat{\overline{y}}_1(\theta))$ and $\widehat{\overline{F}}_\theta(\theta) = 1 - \widehat{F}_{y_1}(\widehat{\underline{y}}_1(\theta))$.

Finally, Theorem 3.2 provides bounds on the differences $S_1(y) - S_1(y')$, which take the following simple form when $y = y_n, y' = y_{n-1}$:

$$\begin{aligned} S_1(y_n) - S_1(y_{n-1}) &\geq t_1(y_n) - t_1(y_{n-1}) + \min_{u\in[y_{n-1},y_n]} q_1(u)\left[\ln\theta_{n-1} - \ln\theta_n\right], \\ S_1(y_n) - S_1(y_{n-1}) &\leq t_1(y_n) - t_1(y_{n-1}) + \max_{u\in[y_{n-1},y_n]} q_1(u)\left[\ln\theta_{n-1} - \ln\theta_n\right], \end{aligned}$$

[26] The number of elements in $\widehat{I}_1$, as well as the range of the sequence $(\widehat{y}_n)_{n\in\widehat{I}_1}$, depends on y_0. Thus, one can choose y_0 so as to maximize one of these two criteria.

where $q_1(.) = t_1'(y)(1 - F_{y_1}(y))/f_{y_1}(y)$. Let $\underline{\Delta S_1}_n$ and $\overline{\Delta S_1}_n$ denote these bounds, we estimate them by

$$\begin{aligned}\widehat{\underline{\Delta S_1}}_n &= t_1(\widehat{y}_n) - t_1(\widehat{y}_{n-1}) + \inf_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} \widehat{q}_1(u)\left[\ln\widehat{\theta}_{n-1} - \ln\widehat{\theta}_n\right],\\ \widehat{\underline{\Delta S_1}}_n &= t_1(\widehat{y}_n) - t_1(\widehat{y}_{n-1}) + \sup_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} \widehat{q}_1(u)\left[\ln\widehat{\theta}_{n-1} - \ln\widehat{\theta}_n\right],\end{aligned}$$

where $\widehat{q}_1(.) = t_1'(y)(1 - \widehat{F}_{y_1}(y))/\widehat{f}_{y_1}(y)$ and $\widehat{f}_{y_1}$ is a kernel estimator of f_{y_1} with bandwidth h_{1N} and kernel function K_1. Bounds on S_2 can be obtained in a same way, using the fact that $\theta_2(.) = (\delta_2/\delta_1)\theta_1(.)$ by the vertical transform.

Theorem 4.1 below establishes the consistency of our estimators, under the following conditions.

Assumption 9 *For all $n \in I_1$, $y_n \in \mathring{\mathcal{Y}}_1$.*

Assumption 10 *The kernel function K_1 is defined on a compact set and admits a bounded derivative on this set. It satisfies $\int K_1(u)du = 1$ and $\int uK_1(u)du = 0$.*

Assumption 9 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 10 is standard in nonparametric estimation and ensures the uniform consistency of the kernel estimator of f_{y_1}.

Theorem 4.1 *Suppose that $K = 2$, $h_{1N} \to 0$, $Nh_{1N}/\ln N \to \infty$ and assumptions 1-5 and 8- 10 hold. Then for all $n \in \mathbb{Z}$, $P(n \in \widehat{I}_1) \to 1_{n\in I_1}$. Moreover, for all $n \in I_1$ as $N \to \infty$,*

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

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

$$(\widehat{\underline{\Delta S_1}}_n, \widehat{\overline{\Delta S_1}}_n) \xrightarrow{\mathbb{P}} (\underline{\Delta S_1}_n, \overline{\Delta S_1}_n)$$

Finally, $\widehat{\underline{C'}}(.)$ and $\widehat{\overline{C'}}(.)$ (resp. $\widehat{\underline{F}}_\theta(.)$ and $\widehat{\overline{F}}_\theta(.)$) are consistent on every $y \in \mathcal{Y}_1$ (resp. $\theta \in \Theta$) except on the sequence $(y_n)_{n\in I_1}$ (resp. $(\theta_n)_{n\in I_1}$).

4.2 Estimation with three or more menus of contracts

When $K \geq 3$, the model is fully identified by Theorem 3.5. However, estimation is rather involved in this case since identification of $y_1(.)$ is achieved by iterating an infinite number of times the process of vertical and horizontal transforms. For a given sample size, one can expect the accuracy of the estimators to become poor as the number of iterations increases. On the other hand, if the number of iteration is too low, interpolation between points badly describes the true function. Hence, as usual in nonparametric estimation, there is a trade-off between bias and variance. Increasing the number of iteration reduces the bias at the cost of a larger variance, and conversely. As the sample size increases, the number of iterations m_N has to tend to infinity in order to make the bias tend to zero, but at an appropriate rate to keep the variance low.

For the ease of notations, we focus on the case where $K = 3$ here. In the following, we show consistency of $C'(.)$ (resp. $F_\theta(.)$) on any compact interval strictly included in $\mathcal{Y}_1$ (resp. in Θ). Because the large support condition 7 must hold on such intervals, we have to strengthen slightly this assumption here.

Assumption 7' *(large support 2) There exists a compact interval $\mathbb{Y}_1 \subsetneq \mathcal{Y}_1$ such that $H_{12}(H_{12}(\mathbb{Y}_1) \cap H_{13}(\mathbb{Y}_1)) \cap H_{13}(H_{12}(\mathbb{Y}_1) \cap H_{13}(\mathbb{Y}_1)) \cap \mathbb{Y}_1 \neq \emptyset$.*

As shown in the proof of Theorem 3.5 and in a similar way to the case $K = 2$, we can identify $\theta_1(.)$ on a sequence $(y_{k,l})_{(k,l)\in I_2}$ (where $I_2 \subset \mathbb{Z}^2$), defined by induction as follows. Firstly, $(0,0) \in I_2$ and $y_{0,0} = y_0$, where $y_0 \in H_{12}(H_{12}(\mathbb{Y}_1) \cap H_{13}(\mathbb{Y}_1)) \cap H_{13}(H_{12}(\mathbb{Y}_1) \cap H_{13}(\mathbb{Y}_1)) \cap \mathbb{Y}_1$. Secondly, if $(k-1, l) \in I_2$ with $k - 1 \geq 0$, and if $H_{12}(y_{k-1,l}) \in \mathcal{Y}_1$, then $(k,l) \in I_2$ and $y_{k,l} = H_{12}(y_{k-1,l})$. Similarly, if $(k, l-1) \in I_2$ with $l - 1 \geq 0$, and if $H_{13}(y_{k,l-1}) \in \mathcal{Y}_1$, then $(k,l) \in I_2$ and $y_{k,l} = H_{13}(y_{k,l-1})$. Negative values can be handled similarly. Note that there is no contradiction in the definition of the sequence, since if $H_{12}(y_{k-1,l}) \in \mathcal{Y}_1$ and $H_{13}(y_{k,l-1}) \in \mathcal{Y}_1$, then $H_{12}(y_{k-1,l}) = H_{13}(y_{k,l-1})$.

Let $Y_1 = [\underline{y}, \overline{y}]$ satisfies $\mathbb{Y}_1 \subsetneq Y_1 \subsetneq \mathcal{Y}_1$ and m_N, the number of iterations, be fixed. We estimate I_2 and $(y_{k,l})_{(k,l)\in I_2}$ by $\widehat{I}_2$ and $(\widehat{y}_{k,l})_{(k,l)\in \widehat{I}_2}$, defined by induction as follows:
- $(0,0) \in \widehat{I}_2$ and $\widehat{y}_{0,0} = y_0$;
- $(k,l) \in \widehat{I}_2$ if and only if $|k| + |l| \leq m_N$, there exists $(k', l') \in \widehat{I}_2$ such that $|k'| + |l'| = |k| + |l| - 1$ and $\widehat{H}_{21}^{k-k'} \circ \widehat{H}_{23}^{l-l'}(\widehat{y}_{k',l'}) \in Y_1$. In this case, we let, when $k, l > 0$ (other cases write similarly),

$$\widehat{y}_{k,l} = \frac{1}{\mathbb{1}_{(k-1,l)\in\widehat{I}_2,\widehat{H}_{12}(\widehat{y}_{k-1,l})\in Y_1}+\mathbb{1}_{(k,l-1)\in\widehat{I}_2,\widehat{H}_{13}(\widehat{y}_{k,l-1})\in Y_1}}\times \left[\widehat{H}_{12}(\widehat{y}_{k-1,l})\mathbb{1}_{(k-1,l)\in\widehat{I}_2,\widehat{H}_{12}(\widehat{y}_{k-1,l})\in Y_1}+\widehat{H}_{13}(\widehat{y}_{k,l-1})\mathbb{1}_{(k,l-1)\in\widehat{I}_2,\widehat{H}_{13}(\widehat{y}_{k,l-1})\in Y_1}\right].$$

Thus, as explained before, our estimator of $(y_{k,l})_{(k,l)\in I_2}$ is defined for small values of $|k|+|l|$ only. When support conditions hold, the estimator of $y_{k,l}$ can be defined by using $y_{k,l} = H_{12}(y_{k-1,l})$ or $y_{k,l} = H_{13}(y_{k,l-1})$. The two corresponding estimators do not coincide in general. In this case, we simply take the mean of these two estimators.[27]

As in the previous subsection, we let $\theta_{k,l} = \theta_1(y_{k,l})$ for all $(k,l)\in I_2$. By the separability assumption on the transfert functions, $\theta_{k,l} = (\delta_1/\delta_2)^k(\delta_1/\delta_3)^l\theta_0$. We simply estimate them by $\widehat{\theta}_{k,l} = (\delta_1/\delta_2)^k(\delta_1/\delta_3)^l\theta_0$, for all $(k,l)\in\widehat{I}_2$.

Now, let us define our estimators of $\theta_1(.)$ and $y_1(.)$. Let $\widehat{\underline{y}} = \min_{(k,l)\in\widehat{I}_2}\widehat{y}_{k,l}$ and define $\widehat{\overline{y}}$ similarly. Then we let $\widehat{\theta}_1(\widehat{y}_{k,l}) = \theta_{k,l}$ for all $(k,l)\in\widehat{I}_2$, and define $\widehat{\theta}_1(.)$ by linear interpolation for other values of $y\in[\widehat{\underline{y}},\widehat{\overline{y}}]$. For all $y\in[\underline{y},\widehat{\underline{y}}]$, we let $\widehat{\theta}_1(y) = \widehat{\theta}_1(\widehat{\underline{y}})$ and define similarly $\widehat{\theta}_1(y)$ for $y\geq\widehat{\overline{y}}$. In a same manner, we let $\widehat{y}_1(\theta_{k,l}) = \widehat{y}_{k,l}$ for all $(k,l)\in\widehat{I}_2$, and extend it on $[\widehat{\underline{\theta}} = \min_{(k,l)\in\widehat{I}_2}\theta_{k,l}, \widehat{\overline{\theta}} = \max_{(k,l)\in\widehat{I}_2}\theta_{k,l}]$ by linear interpolation.[28] Then, for $\theta>0$ and $y\in Y_1$, the estimators of $C'(.)$ and $F_\theta(.)$ are defined by

$$\widehat{C}'(y) = \frac{t_1'(y)}{\widehat{\theta}_1(y)}$$

$$\widehat{F}_\theta(\theta) = 1-\widehat{F}_{y_1}\left(\widehat{y}_1((\theta\wedge\widehat{\overline{\theta}})\vee\widehat{\underline{\theta}})\right)$$

To estimate $S_1(.)$, note that, by an integration by part,

$$S_1(y)-S_1(y_0) = t_1(y)-t_1(y_0)-(q_1(y)\ln\theta_1(y)-q_1(y_0)\ln\theta_1(y_0))+\int_{y_0}^{y} q_1'(u)\ln\theta_1(u)du.$$

where $q_1'(u) = q_1(u)(t''(u)/t'(u)-f_{y_1}'(u)/f_{y_1}(u))-t'(u)$. Let us define $\widehat{f}_{y_1}$ as in the previous subsection and let $\widehat{f}_{y_1}'$ denote the kernel estimator of f_{y_1}' with bandwidth h_{2N} and kernel function K_1. We estimate $\Delta S_1(.) = S_1(.)-S_1(y_0)$ by

$$\widehat{\Delta S}_1(y) = t_1(y)-t_1(y_0)-\left(\widehat{q}_1(y)\ln\widehat{\theta}_1(y)-\widehat{q}_1(y_0)\ln\theta_0\right)+\int_{y_0}^{y}\widehat{q}_1'(u)\ln\widehat{\theta}_1(u)du,$$

where $\widehat{q}_1'(u) = \widehat{q}_1(u)(t''(u)/t'(u)-\widehat{f}_{y_1}'(u)/\widehat{f}_{y_1}(u))-t'(u)$.

[27] We use the weights $(1/2,1/2)$ for the sake of simplicity here. The estimator could be improved by taking weights inversely proportional to estimators of the standard errors of $\widehat{H}_{12}(\widehat{y}_{k-1,l})$ and $\widehat{H}_{13}(\widehat{y}_{k,l-1})$.

[28] Thus, contrarily to $\widehat{\theta}_1(.)$, $\widehat{y}_1(.)$ is defined on a random set. Note also that $\widehat{y}_1(.)$ cannot be defined as the inverse of $\widehat{\theta}_1(.)$, since this function is not necessarily one to one.

Theorem 4.2 *Suppose that $K = 3$, $h_{1N} \to 0$, $Nh_{1N} \to 0$, $h_{2N} \to 0$, $Nh_{2N}^3/\ln N \to \infty$ $m_N \to \infty$ and $m_N = o(\log(N))$ and assumptions 1-6, 7', 8 and 10 hold. Then, for all interval $Y_1 = [\underline{y}, \overline{y}]$ satisfying $\mathbb{Y}_1 \subsetneq Y_1 \subsetneq \mathcal{Y}_1$, all $y \in (\underline{y}, \overline{y})$ and all $\theta \in (\theta_1(\overline{y}), \theta_1(\underline{y}))$, as $N \to \infty$,*

$$\begin{aligned} \widehat{C}'(y) &\xrightarrow{\mathbb{P}} C'(y) \\ \widehat{F}_\theta(\theta) &\xrightarrow{\mathbb{P}} F_\theta(\theta) \\ \widehat{\Delta S}_1(y) &\xrightarrow{\mathbb{P}} \Delta S_1(y_1) \end{aligned}$$

Theorem 4.2 establishes the consistency of our procedure as soon as the number of iterations tends to infinity, but at a slow rate, $m_N = o(\ln N)$. The reason why we require such a slow rate can be basically explained as follows. As mentioned before, and because the estimator is defined by induction, we can expect the accuracy of $\widehat{y}_{k,l}$ to decrease as $|k| + |l|$ increases. Suppose that the mean square error (MSE) of $\widehat{y}_{k,l}$ is multiplied by at most $\lambda > 1$ when $|k| + |l|$ increases. After m_N iterations, the MSE of $\widehat{y}_{k,l}$ can be as large as $\lambda^{m_N - 1} \times M_N$, where M_N denotes the MSE of, say, $\widehat{y}_{1,0} = \widehat{H}_{12}(y_0)$. Since one can show that $M_N = O(1/N)$, letting $m_N = o(\ln N)$ ensures that $\lambda^{m_N-1} \times M_N$ tends to zero.

5 Monte Carlo simulations

In this section, we investigate the finite sample properties of our estimators when $K = 3$. We consider a nonlinear pricing setting à la Maskin and Riley (1984). We suppose to observe three periods ($j = 1, 2, 3$) where a monopoly produces the quantity y at a cost $C_j(y) = c_j y^2/2$. The changes may typically come from variation in the price of an input. The utility of agents satisfies $U = \theta H(y) - t$. The consumers' heterogeneity component θ is distributed according to a Pareto distribution $F_\theta(\theta) = 1 - \theta^{-4}$ on $[1, \infty[$ and we let $H(y) = y$. These functions remain the same during the three periods.

The principal's first order condition satisfies

$$C_j'(y) = H'(y)\left[\theta_j(y) - \frac{1 - F_\theta}{f_\theta}(\theta_j(y))\right]$$

With our specification, we obtain :

$$\theta_j(y) = t_j^{*\prime}(y) = \frac{4c_j}{3} y.$$

In our simulations, we chose $c_1 = 1$, $c_2 = 0.8$ and $c_3 = 1.2$. For each contract j, we drew a sample $(y_{ij} = 3\theta_{ij}/4c_j)_{i=1...N}$ of quantities chosen by $N = 1,000$ agents and replicated our estimation procedure 500 times. We chose a number of iterations $m_N = 3$, a starting point (θ_0, y_0) such that $\theta_0 = F_\theta^{-1}(0.5) \simeq 1.19$ and $y_0 = 3\theta_0/4 \simeq 0.89$, and an interval of estimation $Y_1 = [0.8, 1.6]$. This allows us to recover from 10 to 15 points, depending on the simulation.[29]

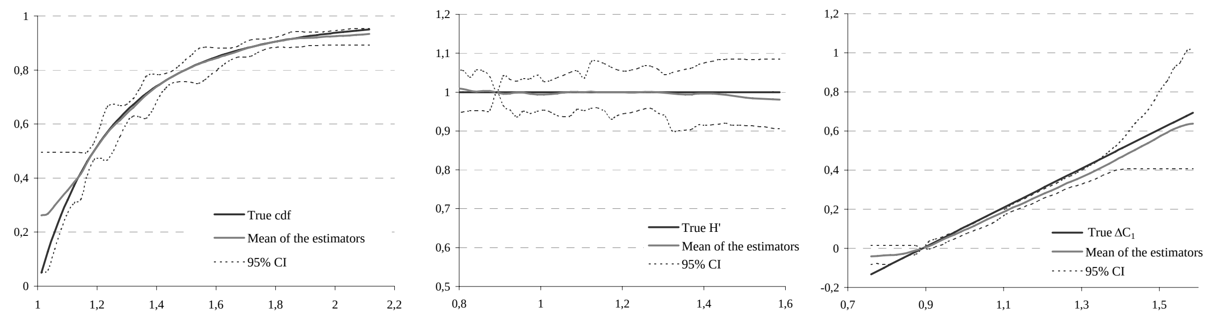

Figure 4: Estimation of $F_\theta(.)$, $H'(.)$ and $\Delta C_1(.) = C_1(.) - C_1(y_0)$.

Figure 4 shows that we are able to recover well the functions of interest. We observe an overestimation of F_θ for small values of θ, which stems from the fact that the estimator of F_θ is constant for all θ inferior to $\min_{(k,l)\in\widehat{I}_2} \theta_{k,l}$. The behavior of $\widehat{H'}$ is satisfactory on the whole interval Y_1. Note that by construction, $\widehat{H'}(y_0)$ the estimator is constant and equal to $H'(y_0)$. The bias of the estimator of ΔC_1 for the principal is comparable to its standard deviation for all $y \leq 1.3$, as usual with estimators involving nonparametric estimation of a density and its derivatives. The variance of the estimator is large for $y \geq 1.3$, as expected since the nonparametric estimator of f'_y/f_y is unstable for large values of y where the denominator f_y is small. Overall, the Monte Carlo simulations show that our recursive procedure performs well. We obtain small mean square errors on the functions of interest, even with a small number of iterations ($m_N = 3$).

[29] We chose a bandwidth of respectively $3\sigma/N^{1/5}$ and $6\sigma/N^{1/5}$, where σ denotes the standard deviation of y, for the kernel estimators of the density of y and its derivative.

6 Application

In this section, we apply our identification and estimation results with two different menus of contracts ($K = 2$) to data between the French national institute of economics and statistics (Insee) and the interviewers it hires to conduct household surveys. For each survey, Insee's interviewers receive a random sample of households close to their residence and have to fulfill a maximum number of face to face interviews. Because Insee cannot compel its interviewers to obtain a given number of interviews, it provides them with incentives through a specific payment scheme. More precisely, each interviewer receives a basic wage plus a bonus for each interview he achieves.

We have 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. For all these households, the basic wage was around 4.6 euros, whereas the bonus for achieving an interview was nearly the same in 2001 and 2002 (resp. 20.3 and 20.2 euros in 2002 euros) but increased to 22.9 euros in 2003. The idea of the application is to use the change between 2001-2002 and 2003 to recover bounds on the cost function and the distribution of the types of the interviewers. Indeed, and as explained below, we believe that this change is exogenous, something necessary for our method to apply. Eventually, assuming that Insee is not optimal and maximizes its objective function only among the class of two part tariffs, we estimate the loss of using such simple contracts instead of the optimal ones.

Our application contributes to the literature of provisions of incentive in 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 observed contracts. Our approach is structural and similar to Ferrall and Shearer (1999) and Paarsch and Shearer (2000), but nonparametric.

6.1 The model

6.1.1 The agent

In a survey j, an interviewer $i \in \{1, ..., N_j\}$ (the agent) receives a random sample of housings to interview. For that purpose, he has to contact the households and then to convince them to accept the interview. The cost of an interview is first due to the contacting process: the interviewer usually has to go to the housing several times, let phone messages, deal with interphones or digicodes... in order to reach the household. It is not unusual for interviewers to go more than five times to the households' houses, which can be more than 20kms far from their owns. Even if reimbursed for the gasoline and the meals, this cost is important as trying to contact the households takes several days. Once contacted, the households can still accept or refuse the questionnaire. This step is important but not really costly for the interviewer. Indeed, trying to persuade the household is far less time consuming than trying to contact it. The last aspect of the cost corresponds to the interview itself. For the surveys considered here, it takes around one hour to go through all the questions, and thus can again be considered negligible compared to the first cost. Discussions with interviewers confirm that the essential part of their job consists of contacting people, and we only model this aspect, taking into account the acceptance rate. Concerning this last point, we have shown in an internal study (see D'Haultfoeuille and Février, 2002) that the acceptance rate does not depend on most of the observable characteristics of the households and of the interviewers. Hence, we suppose that it is independent of i and denote it by a_j.

The other aspect of the agent's model is heterogeneity. In our application, there are two sources of heterogeneity. The first is the standard one, which captures differences between interviewers themselves. The second is related to the sample of households that interviewers have to contact. Indeed, households may be easy or difficult to contact, depending on their characteristics. For instance, it is difficult to contact a young household in a urban area, because it is often absent during the day and because of the digicode or interphone at the entry. Elder people in the countryside, on the contrary, are easier to contact because more at home. We summarize the two aspects of heterogeneity by a parameter $\theta_{ij} \in \mathbb{R}^+$ and denote by $F_{\theta,j}$ the corresponding cumulative distribution function. Finally, the interviewers work for Insee in the same area and usually for a very long time (the average experience was around 6 years in 2003). Thus, they have a very strong knowledge of the difficulties they will face. Furthermore, before trying to contact the households, interviewers must locate precisely the housings of their sample (in order to identify unoccupied or destroyed housings, for instance). Hence, it is reasonable to believe that at the end of this phase, the interviewers have learned the difficulty of their sample and thus know θ_{ij}.

To sum up, an interviewer has to contact a proportion y/a_j of households to obtain a response rate of y on his sample. We assume that the corresponding cost is separable and satisfies $\theta_{ij}C(y/a_j)$. C is independent of j as, given θ_{ij}, it represents the time spent to try to contact the households either by phone or by ringing at their door. There is no obvious reason why the cost should have changed from one year to another as the surveys were drawn in the same way, conducted during the same period and had identical rules for the fieldwork.

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:

$$\max_y \delta_j y - \theta_{ij} C \left(\frac{y}{a_j} \right).$$

6.1.2 The principal

Insee is a public institute which maximizes the social value of each survey. The surveys are used by politicians, researchers... and may differ in the "price" of the information that can be recovered from it. For this reason, we believe that Insee's objective function $S_j(.)$ may differ from one survey j to another.

Besides, we have reasons to believe that Insee's contracts are inoptimal. This assumption is confirmed by two facts. The first is the violation of the Informativeness Principle which states that all factors correlated with performance should be included in the contracts (Prendergast, 1999). Here, for instance, the bonus does not depend 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%). The second is the fact that Insee uses two part tariffs for all its household surveys, not only the PCV ones. This feature seems too peculiar to assume that Insee maximizes his objective function among all contracts. We believe instead that it maximizes his objective function only in the class of two part tariffs. The optimality of the principal is however not required to identify partially C and $F_{\theta,j}$ when $K = 2$.

6.2 An exogeneous change

In 2003, the bonus raised from 20.2 euros to 22.9 euros. We believe that this change stems from a modification of Insee's objective function. 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. We believe that the 2003 survey on education was considered by Insee to be more important than the other ones. Indeed, there was 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 which might explain why Insee decided to increase the bonus.

To apply our method, this change in the 2003 contract has to be exogenous. This means in particular that the distribution functions $F_{\theta,j}(.)$ and the acceptance rates a_j should be the same in all surveys.

As explained above, $F_{\theta,j}(.)$ depends both on the heterogeneity of interviewers and of the samples of households they receive. These samples are drawn from the whole population in France by a program defined at the central level of Insee in Paris and the distribution of the samples' types is thus defined by this procedure. In particular, the sampling designs of the three surveys of 2001, 2002 and 2003 were the same, which means that the heterogeneity stemming from households has the same distribution in the three surveys. Concerning the other side of the heterogeneity, it is possible that selection effects arise following the increase of the bonus in 2003. To recover safely C', and as explained previously, we restrict ourselves in a first step to the stayers, i.e. the interviewers that participate in the three surveys. We come back to the issue of selection effects in subsection 6.4.2.

As for the acceptance rates, they are rather constant over time, around 95%, in the PCV surveys (Le Lan, 2008). These rates are high because these surveys are mandatory and done by a public institute. They do not vary much over time because the willingness to participate in a survey is mainly related with the time households have at their disposal (Le Lan, 2008). The topic of the survey does not seem to play a crucial role in the participation decision. This is reinforced by the fact that the questionnaires of PCV surveys contains a fixed part, always identical for all October editions, which represents more than half of the survey. Finally, we can perform an indirect test of the assumption $a_1 = a_2 = a_3$ by comparing the outcome in 2001 and 2002 for which the paiement rule was similar but topics differ. The fact that nothing changes (see figure 5)[30] should be seen as a weak evidence of independence between the acceptance rate and the topic of the PCV survey. By abuse of notations, we denote $C(./a)$ by $C(.)$ subsequently.

For all these reasons and from our own experience,[31] it seems reasonable to us to assume that the 2003 change is exogenous in the sense that it came from a change in the objective function of the principal rather than from a change in the characteristics of the agents.

[30] The p-value for the two-sided Kolmogorov-Smirnov test is 0.69.

[31] We both worked at Insee in the household survey methodology unit between 2000 and 2003.

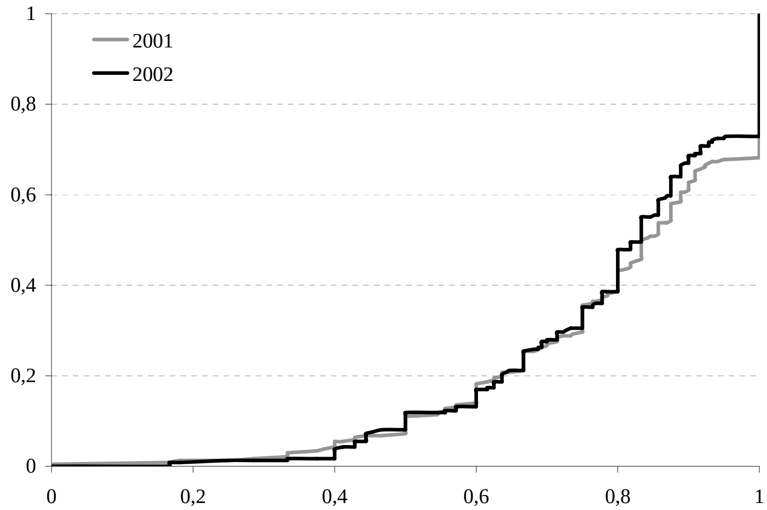

Figure 5: Distribution functions of the response rates for the 2001 and 2002 surveys.

6.3 Descriptive analysis

As already said, we use data on three household living conditions surveys which took place in October 2001, 2002 and 2003 ($j = 1, 2, 3$). Almost all households were eligible to the survey in 2001 and 2002 whereas only families were eligible in 2003. Hence, we restrict our attention in all surveys 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.

Year	Number of Interviewers	Bonus (in 2002 euros)	Average response rate All interviewers	Stayers
2001	377	20.3	79.0%	79.6%
2002	471	20.2	79.8%	80.4%
2003	453	22.9	83.1%	83.8%

Table 1: Descriptive statistics.

Our data consists of 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. As already mentioned, the 2001 and 2002 surveys display very similar patterns.[32] We aggregate them in the rest of the application to obtain more precise results. Besides, selection effects do not seem to be really an issue here. Indeed, the interviewers who did the 2003 survey only obtained an average response rate slightly below the stayers (82.4% versus 83.8%), which is not compatible with the expected effect. We come back to this issue in subsection 6.4.2.

Figure 6 displays the distribution function of the response rates for the 2001-2002 and 2003 surveys. As predicted by the theory,[33] the distribution function of the 2003 survey stochastically dominates the one of 2001-2002, which proves that interviewers react to incentives.[34] We find that on average production increases by 5% when the piece rate increases by 13.4%. This result suggests a significant incentive effect, in line with the results of, e.g., Lazear (2000) and Paarsch and Shearer (2000). Figure 6 also displays several jumps in the distribution functions. The mass points are due to finite approximations of the response rates, and we neglect these error terms in our estimation.[35]

6.4 Results

6.4.1 Nonparametric estimation

We now apply the estimation method corresponding to $K = 2$ to our contract data. Firstly, starting from a middle point $y_0 = 0.6$ (with $\theta_0 = 1$), we estimate $(y_n, \theta_n, F_\theta(\theta_n), C'(y_n))_{n \in I_1}$ as indicated and obtained 12 distinct points which correspond to $n \in \{-3, ..., 8\}$. Figure 7 displays the estimates of the bounds on $F_\theta(.)$ and $C'(.)$, 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

[32]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.

[33]Indeed, equations (3.1) and (3.2) imply that for all y, $t'_i(y) < t'_j(y) \Leftrightarrow F_{y_i}(y) > F_{y_j}(y)$.

[34]The one-sided Kolmogorov-Smirnov test which tests the equality of the distribution functions rejects the null hypothesis at 5%.

[35]Similarly, in theory, $y \in [0, a]$. However, because of the finite number of interviews, $y \in [0, 1]$ empirically.

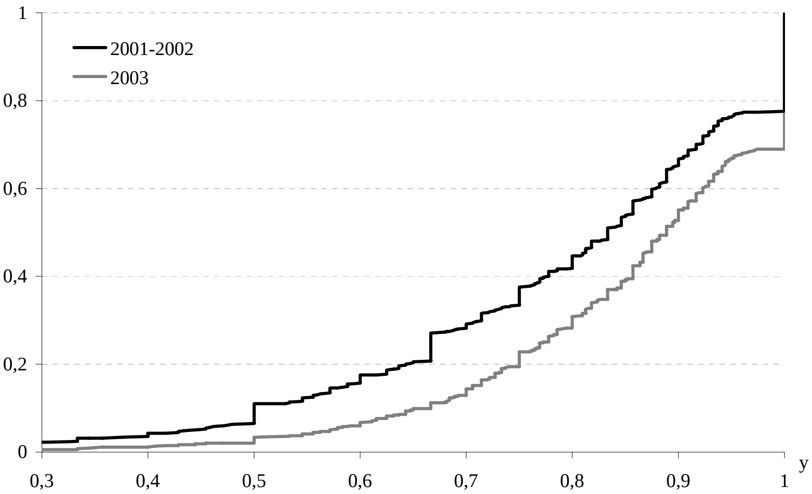

Figure 6: Distribution functions of the response rates for 2001-2002 and 2003 surveys.

significantly 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. Finally, the width of the confidence intervals on the bounds of F_θ (resp. C') increases with $|\theta - 1|$ (resp. $|y - 0.6|$), reflecting the fact that, as expected, the estimation error increases with $|n|$.

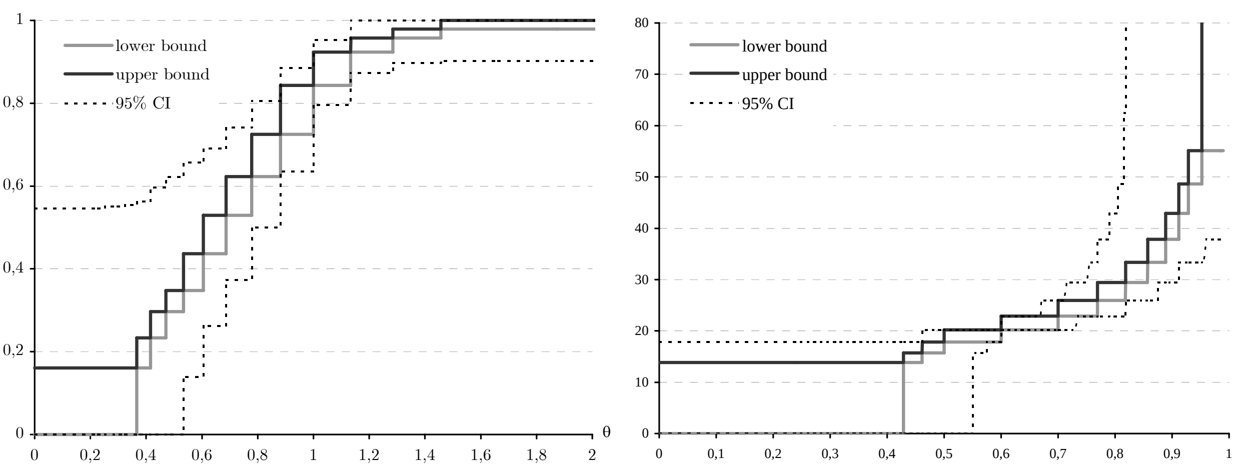

Figure 7: Estimated bounds on $F_\theta(.)$ and $C'(.)$.

6.4.2 Selection effects

Before turning to the cost of using two part tariffs, we check whether selection effects are important in the 2003 survey. To investigate this issue, we compare the estimates of the distribution of θ for the stayers, the interviewers who did the 2001 or 2002 survey only (the 2001-2002 movers hereafter) and those who participate in the 2003 survey only (the 2003 movers hereafter).

Figure 8 displays the estimates of $(F_\theta(\theta_n))_{n\in I_1}$ for the three populations, and the 95% confidence interval of the distribution of the stayers' type. The estimated sequences of the movers remain within the confidence band, except at one point. Moreover, the distribution of the 2003 movers' types seems to stochastically dominate the one of the stayers, in contradiction with the prediction that the 2003 survey should attract better interviewers. Thus, we reject selection effects in our application. The rotation pattern observed between the interviewers in the different surveys seems to be exogenous and not related to the payment scheme.

The fact that the distribution of movers' types does not significantly differ from the stayers' one also indicate that the estimated bounds on F_θ we have obtained are consistent estimates of the ones on the whole population of interviewers, and not only of the stayers. Thus, these estimates can be used for analyzing the principal's program, to which we now turn.

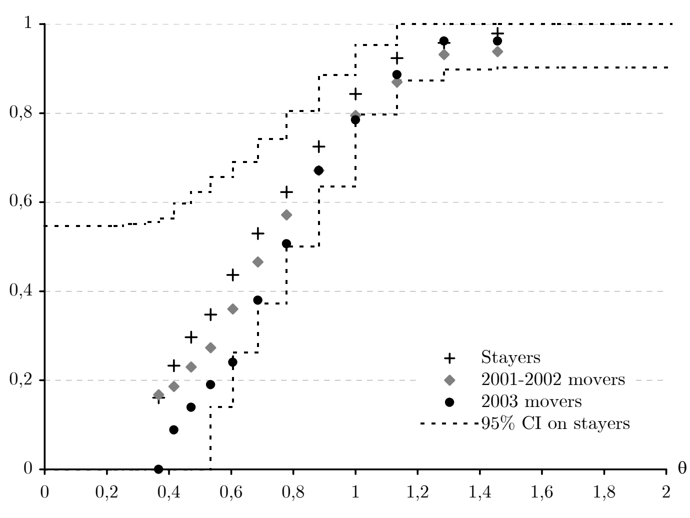

Figure 8: Estimation of $F_\theta(\theta_n)$ for the three populations of interviewers.

6.4.3 The cost of using inefficient contracts

As previously mentioned, we believe that Insee does not implement the optimal contract but rather chooses the optimal linear one. We now use the previous estimates to measure the cost of using such simple but inefficient incentives. Because of inoptimality, equation (3.3) is not valid here. It can be shown in this case that the model is not nonparametrically identified. Thus, we adopt a parametric approach and impose a structure consistent with our previous nonparametric analysis. More precisely, we suppose 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 $\ln[(1-y)/y]$.[36] We obtain $\widehat{a} = 1.87\,(0.02)$, $\widehat{b} = 2.21\,(0.08)$, $\widehat{\alpha} = 17.2\,(0.55)$ and $\widehat{\beta} = 0.39\,(0.01)$. Figure 9 shows that these parametric forms perfectly fit the nonparametric estimated points. Finally, we suppose that Insee's objective function takes a linear form, $S_j(y) = \lambda_j y$. λ_j represents the "price" of the information contained in a household's answers, i.e. the social value of an interview in survey j. In our framework, $\lambda_1 = \lambda_2 < \lambda_3$ as Insee values more the 2003 answers than those of 2001 or 2002. Under these parametric restrictions, we are able to estimate λ_3. We find the social value of an interview to be $\widehat{\lambda}_3 = 88.6$ euros.

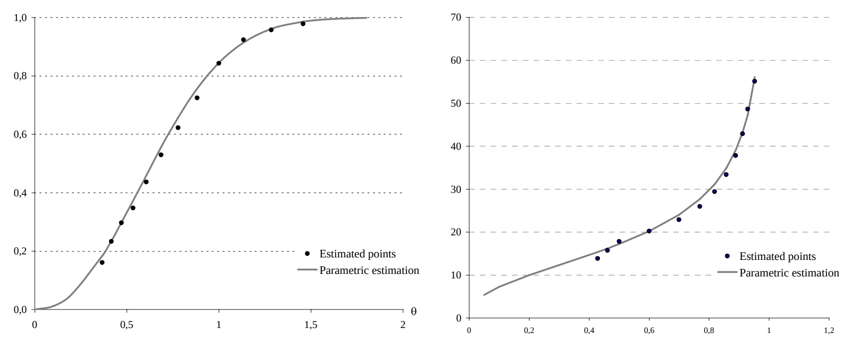

Figure 9: Parametric estimation of F_θ and C'.

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 two part tariffs is around 10% and that the response rate decreases by 8% compared to optimal contracts. This result

[36] Further details on this estimation and on this subsection in general can be found in the supplementary material.

contrasts with the idea that simple contracts can be quite inefficient. Ferrall and Shearer (1999), for instance, evaluate the loss of using such simple contracts to be around 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 two-part tariff contracts compared to nonlinear ones: they are less costly to implement and almost efficient. A recent empirical paper of Miravete (2007), which reports a loss of only 3%, also supports this claim. These two results are in line with the theoretical findings of Wilson (1993, section 6.4), Rogerson (2003) and Chu and Sappington (2007), who show that simple tariffs secure at least 89%, 75% and 74% of the maximal surplus, respectively. Studying auctions, Neeman (2003) also proves that simple English auctions generates an expected price that is more than 80% of the value of the object to the bidder with the highest valuation.

Environment	Pay method	E[surplus]	Relative	E[response rate]
Full information	Optimal contract	72.9	1.00	0.99
Incomplete information	Optimal contract	60.8	0.83	0.92
Incomplete information	Linear contract	55.0	0.75	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 and with the two part tariff is 75% 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 tarification.

7 Conclusion

This work contributes to the recent structural analysis of incentive problems.[37] The model we study encompasses the basic regulation, nonlinear pricing and price discrimination models, first price auctions and simple insurance settings. We show in particular that when contracts vary through an instrument, one can obtain partial or full nonparametric identification of these models. Our results are based on a new recursive method that we apply both for identification and estimation. We believe that this method should be useful

[37]For a structural analysis of moral hazard, see Ke (2008).

for analyzing identification of adverse selection models with discrete choices and of more complex versions of the model. Although testability is not considered here, we also believe that exogenous changes are important to test the optimality of contracts or the presence of asymmetries.

8 Appendix: proofs

Proof of Theorem 3.1

For any strictly decreasing and differentiable function $\theta_1(.)$, it is possible to define $F_\theta(.)$, $C'(.)$, and S_1' using respectively equation (3.1), (3.2) and (3.3). For the model not to be identified, it suffices to prove that two sets of such functions satisfy assumptions 1 and 3. Because these assumptions are satisfied for the true function $\theta_1^0(.)$, the assumption is satisfied also locally around $\theta_1^0(.)$. Hence, the model is not identified.

If $F_\theta(.)$ (resp. $C'(.)$) is known, equation (3.1) (resp. (3.2)) enables to identify $\theta_1(.)$ on $\mathcal{Y}_1$. Then $C'(.)$ (resp. $F_\theta(.)$) is identified on Θ (resp. $\mathcal{Y}_1$) using the previous equations. Finally, $S_1'(.)$ is identified by (3.3).

Finally, if S_1' is known, (3.3) is equivalent to

$$\ln(\theta_1(y)) = \ln(\theta_0) + \int_{y_0}^{y} \left(1 - \frac{S_1'(u)}{t_1'(u)}\right) \frac{f_{y_1}(u)}{1 - F_{y_1}(u)} du \tag{8.1}$$

Hence $\theta_1(.)$ is identified. By (3.1) and (3.2) $F_\theta(.)$ and $C'(.)$ are also identified. ■

Proof of Theorem 3.2

We first prove part 1 of the theorem. Because $H_{12}(.)$ and $H_{21}(.)$ are identified, it follows from the discussion before Theorem 3.2 that $\theta_1(.)$ is point identified on $(y_n)_{n\in\mathbb{Z}}$. For other y, we get, by monotonicity of $\theta_1(.)$,

$$\sup_{n:\, y_n \geq y} \theta_1(y_n) = \underline{\theta_1}(y) \leq \theta_1(y) \leq \overline{\theta_1}(y) = \inf_{n:\, y_n \leq y} \theta_1(y_n)$$

where the supremum (resp. the infimum) is set to zero (resp. infinity) when the set is empty. Similarly, $\underline{y}(\theta) \leq y_1(\theta) \leq \overline{y}(\theta)$. Then equations (3.1) and (3.2) imply inequalities (3.6) and (3.7).

Now let us turn to part 2. Under optimality of the contracts, we have, for $k \in \{1, 2\}$ and all $y > y'$, by (3.3):

$$S_k(y) - S_k(y') = t_k(y) - t_k(y') + \int_{y'}^{y} q_k(u) \frac{-\theta_k'(u)}{\theta_k(u)} du.$$

Because $-\theta_k'(u)/\theta_k(u) > 0$ for all u,

$$\begin{aligned} S_k(y) - S_k(y') &\leq t_k(y) - t_k(y') + \max_{u\in[y,y']} q_k(u) \left[\ln \theta_k(y') - \ln \theta_k(y)\right]. \\ &\leq t_k(y) - t_k(y') + \max_{\in[y,y']} q_k(u) \left[\ln \overline{\theta}_k(y') - \ln \underline{\theta}_k(y)\right]. \end{aligned}$$

The lower bound can be obtained similarly. ■

Proof of Theorem 3.3

By the normalization and the fact that $y_0 \in \overset{\circ}{\mathcal{Y}}_1$, we can always fix $0 < \theta_0 < \infty$ such that $\theta_0 = \theta_1(y_0)$. We also suppose, without loss of generality that $t_2'(.) > t_1'(.)$ for all $y < y_0$.

Let $y_\alpha < y_0$, and define the increasing sequence $(y^n)_{n\in\mathbb{N}}$ by $y^0 = y_\alpha$ and, for all $n \geq 1$,

$$y^n = H_{12}^n(y_\alpha)\mathbb{1}_{H_{12}(y^{n-1})\in\mathcal{Y}_1} + y^{n-1}\mathbb{1}_{H_{12}(y^{n-1})\notin\mathcal{Y}_1}.$$

We see that $y^n < y_0$ for all $n \in \mathbb{N}$. Indeed, the result is true for $n = 0$. Moreover, if it holds for $n - 1$, then $y^n = H_{12}(y^{n-1}) < H_{12}(y_0) = y_0$ 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_0, so that it admits a limit y^∞ which satisfies $y^\infty = H_{12}(y^\infty)$ and $y^\infty = y_0$. Hence $y^\infty = y_0$.

Now, by the first order condition,

$$\theta_1(y^{n+1}) = V_{21}(\theta_1(y^n), y^{n+1}) = \frac{t_1'(y^{n+1})}{t_2'(y^{n+1})}\theta_1(y^n).$$

Thus, by a straightforward induction,

$$\theta_1(y^n) = \prod_{i=1}^{n} \left[\frac{t_1'(y^i)}{t_2'(y^i)}\right] \theta_1(y_\alpha).$$

Because $(y^n)_{n\in\mathbb{N}}$ converges to y_0 and $\theta_1(.)$ is continuous, the sequence $(\theta_1(y^n))_{n\in\mathbb{N}}$ converges to θ_0. Hence,

$$\theta_0 = \prod_{i=1}^{\infty} \left[\frac{t_1'(y^i)}{t_2'(y^i)}\right] \theta_1(y_\alpha).$$

Because $0 < \theta_0 < \infty$, the product in the right hand side is strictly positive and finite and we have

$$\theta_1(y_\alpha) = \frac{\theta_0}{\prod_{i=1}^{\infty}\left[\frac{t_1'(y^i)}{t_2'(y^i)}\right]}.$$

Because the right term can be recovered from the data, this proves that $\theta_1(y_\alpha)$ is identified. $y_\alpha < y_0$ was arbitrary, so that $\theta_1(.)$ is identified on $\mathcal{Y}_1 \cap]-\infty, y_0]$.

If $M = 0$, i.e. there is a unique crossing point, we can identify $\theta_1(y_\alpha)$ for all $y_\alpha > y_0$ as previously, using a decreasing sequence instead of an increasing one. Thus $\theta_1(.)$ is actually identified on $\mathcal{Y}_1$ in this case. If $M \geq 1$, we can identify similarly $\theta_1(y_\alpha)$ for all $y_1 > y_\alpha > y_0$. $\theta_1(y_1)$ is then identified by continuity. Hence, $\theta_1(.)$ is identified on $\mathcal{Y}_1 \cap]-\infty, y_1]$. A straightforward induction on M then shows that $\theta_1(.)$ is identified on the whole set $\mathcal{Y}_1$.

Finally, by Theorem 3.1, F_θ is identified on Θ and C' and $S_1'(.)$ are identified on $\mathcal{Y}_1$. Furthermore, by the horizontal transformation, $\theta_2(.)$ is also identified on $\mathcal{Y}_2$. Hence, $S_2'(.)$ is identified on $\mathcal{Y}_2$, and C' is identified on $\mathcal{Y}_1 \cup \mathcal{Y}_2$. ■

Proof of Theorem 3.4

$\widetilde{\mathbb{Y}}$ is the set of $\cup_{i=1,..,K}\mathcal{Y}_i$ such that the functions $\theta_i(.)$ 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 $(y, \theta_i(y))$ is identified.

- For all j, $H_{ij}(y)$ and $\theta_j(H_{ij}(y)) = \theta_i(y)$ are known because H_{ij} is identified. Hence the points $(H_{ij}(y), \theta_j(H_{ij}(y)))$ are identified.
- For all j such that $y \in \mathcal{Y}_j$, $\theta_j(y) = V_{ij}(\theta_i(y), y)$ is known because V_{ij} is identified. Hence the points $(y, \theta_j(y))$ 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 theorem. 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 $\theta_i(.)$, we get $\theta_i(y) = \lim_{n\to\infty} \theta_i(y_n)$. Hence, $\theta_i(.)$ is actually identified on $\overline{\mathbb{Y}} \cap \mathcal{Y}_i$. ■

Proof of Theorem 3.5

By assumption 7, there exists $y_0 \in \mathcal{Y}_2 \cap \mathcal{Y}_3$ such that $H_{13}(y_0) \in \mathcal{Y}_1$ and $H_{12}(y_0) \in \mathcal{Y}_1$. Moreover, as $\mathcal{Y}_2 \cap \mathcal{Y}_3 \subset \mathcal{Y}_1$, $y_0 \in \mathcal{Y}_1$. As usual, let θ_0 satisfy $\theta_1(y_0) = \theta_0$. Now, let us define $\widetilde{\Theta} = \theta_1(\mathbb{Y} \cap \mathcal{Y}_1)$, $E_2 = \delta_2/\delta_1 < 1$, $E_3 = \delta_3/\delta_1 > 1$ and

$$A_n = \{E_2^p E_3^q \theta_0,\ |p| + |q| = n\} \cap \Theta.$$

Suppose that $A_n \subset \widetilde{\Theta}$ for all $n \geq 0$. Then $\cup_n A_n \subset \widetilde{\Theta}$. By lemma 3.1 in the supplementary material and assumption 6, $\cup_n A_n \subset \widetilde{\Theta}$ is dense in Θ. By continuity of $y_1(.)$, we get

$$\mathcal{Y}_1 = y_1(\Theta) = y_1(\overline{\widetilde{\Theta}}) \subset \overline{y_1(\widetilde{\Theta})} = \overline{\mathbb{Y}} \cap \mathcal{Y}_1.$$

Thus, $\mathcal{Y}_1 \subset \overline{\mathbb{Y}}$ and by Theorem 3.4, $\theta_1(.)$ is identified on $\mathcal{Y}_1$. Reasoning as in Theorem 3.3, this proves the identification of the functions of interest.

Hence, we just have to show that $A_n \subset \widetilde{\Theta}$. We proceed by induction on n. The result is trivial for $n = 0$. Now, suppose that it is true for n and let $\theta = E_2^p E_3^q \theta_0 \in A_{n+1}$. We have to show that $\theta \in \widetilde{\Theta}$.

First, suppose that $\theta \geq \theta_0$ and $p < 0$. The vertical transform $V_{12}(., y_0)$ is well defined because $y_0 \in \mathcal{Y}_1 \cap \mathcal{Y}_2$. Moreover, by assumption 5, $\theta_2(y_0) = V_{12}(\theta_0, y_0) = E_2\theta_0$. Thus, $E_2\theta_0 \in \Theta$. Hence, $\overline{\theta} \geq \theta > E_2\theta \geq E_2\theta_0 \geq \underline{\theta}$, so that $E_2\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$. Moreover,

$$E_2\theta = E_2^{p+1} E_3^q \theta_0, \quad |p+1| + |q| = (-p-1) + |q| = n.$$

Thus $E_2\theta \in A_n \subset \widetilde{\Theta}$ by the induction hypothesis.

Now, by definition of $\widetilde{\Theta}$, there exists $y \in \mathbb{Y} \cap \mathcal{Y}_1$ such that $\theta_1(y) = E_2\theta$. To show that $\theta \in \widetilde{\Theta}$, we establish that $H_{12}(y) \in \mathbb{Y} \cap \mathcal{Y}_1$ and $\theta = \theta_1(H_{12}(y))$. By definition of $\mathbb{Y}$, $H_{12}(y) \in \mathbb{Y}$. Besides, $y_0 \in \mathcal{Y}_1 \cap \mathcal{Y}_2$ and $\theta_1(y_0) \leq \theta$. Furthermore, because $\theta_1(.) > \theta_2(.)$, we have $\sup \theta_1(\mathcal{Y}_1 \cap \mathcal{Y}_2) = \overline{\theta} \geq \theta$. Thus, there exists $\widetilde{y} \in \mathcal{Y}_1 \cap \mathcal{Y}_2$ such that $\theta_1(\widetilde{y}) = \theta$. Because $\widetilde{y} \in \mathcal{Y}_1 \cap \mathcal{Y}_2$, the vertical transform $V_{12}(., \widetilde{y})$ is well defined and $V_{12}(\theta, \widetilde{y}) = E_2\theta$. Moreover, $\theta_2(H_{12}(y)) = E_2\theta$ by definition of $H_{12}(.)$. Thus, $H_{12}(y) = \widetilde{y} \in \mathcal{Y}_1$ and $\theta = \theta_1(H_{12}(y))$. This proves that $\theta \in \widetilde{\Theta}$.

The proof is similar with $\theta \geq \theta_0$, $p \geq 0$ and $q > 0$[38] using $E_3^{-1}\theta$ instead of $E_2\theta$. The case $\theta < \theta_0$ can be handled similarly using $E_2^{-1}\theta$ and $E_3\theta$.

Eventually, we get $A_{n+1} \subset \widetilde{\Theta}$, so that the result is true for all n, which concludes the proof. ■

[38]Note that the case where $p \geq 0$ and $q \leq 0$ is impossible when $\theta \geq \theta_0$.

Proof of Theorem 3.6

By uniformity of θ, (3.1) now writes as $\theta(y,z) = 1 - F_{y|z}(y|z)$. Thus, $\theta(.,z)$ is identified on $\mathcal{Y}_z = \{y : \exists \theta \in \Theta : \theta(y,z) = \theta\}$. By (3.2), $\frac{\partial C}{\partial y}(.,.)$ is also identified on $\{(y, \theta(y,z)), y \in \mathcal{Y}_z\}$, for all $z \in \mathcal{Z}$.

Now, for all $y \in \mathcal{Y}_z$,

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

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

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

Some manipulations show that the first order condition of the principal satisfies here

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

The function $\frac{\partial S}{\partial y}(.,z)$ is therefore identified on $\mathcal{Y}_z$. ■

Proof of Theorem 4.1

1. $P(n \in \widehat{I}_1) \to 1_{n \in I_1}$ for all $n \in \mathbb{Z}$, and consistency of $(\widehat{y}_n, \widehat{C'(y_n)}$ and $(\widehat{\theta}_n, \widehat{F_\theta(\theta_n)})$, for all $n \in I_1$.

We first prove that for all $n \in I_1$ and for all $\varepsilon > 0$, as $N \to \infty$,

$$P(n \in \widehat{I}_1, |\widehat{y}_n - y_n| \le \varepsilon) \to 1 \tag{8.2}$$

Let us proceed by induction on n. The proposition is true when $n = 0$. Suppose that it holds for $n - 1 > 0$ and let us prove that it holds for n (the proof is similar for negative values). First, let us prove that for all $\varepsilon > 0$,

$$P(n - 1 \in \widehat{I}_1, |\widehat{H}_{12}(\widehat{y}_{n-1}) - y_n| \le \varepsilon) \to 1 \tag{8.3}$$

Without loss of generality, we can focus only on $\varepsilon > 0$ such that $B(y_{n-1}, \varepsilon) \subset \overset{\circ}{\mathcal{Y}}_1$, where $B(x,r)$ is the closed ball of center x and radius r. Such $\varepsilon > 0$ exist by assumption 9. By lemma 3.2 in the supplementary material, $M = 1 \vee \sup_{x \in B(y_{n-1}, \varepsilon)} |H'_{12}|(x) < \infty$. Moreover, by the induction hypothesis and lemma 3.3 in the supplementary material, for all N large enough, the event

$$E_0 = \left\{ n-1 \in \widehat{I}_1, |\widehat{y}_{n-1} - y_{n-1}| < \varepsilon/2M, \sup_{x\in B(y_{n-1},\varepsilon)} |\widehat{H}_{12}(x) - H_{12}(x)| < \varepsilon/2 \right\}$$

holds with an arbitrarily large probability. Now, under E_0,

$$\begin{aligned} |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| &\leq |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(\widehat{y}_{n-1})| + |H_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \\ &\leq \sup_{x\in B(y_{n-1},\varepsilon)} |\widehat{H}_{12}(x) - H_{12}(x)| + M|\widehat{y}_{n-1} - y_{n-1}| \\ &\leq \varepsilon. \end{aligned}$$

This proves (8.3) and . Now,

$$\begin{aligned} P(n \in \widehat{I}_1, |\widehat{y}_n - y_n| \leq \varepsilon) &= P\left(n \in \widehat{I}_1, n-1 \in \widehat{I}_1, |\widehat{H}_{12}(\widehat{y}_{n-1}) - y_n| \leq \varepsilon\right) \\ &\geq P\left(n-1 \in \widehat{I}_1, |\widehat{H}_{12}(\widehat{y}_{n-1}) - y_n| \leq \varepsilon\right) - P\left(n \notin \widehat{I}_1\right) \end{aligned}$$

By (8.3) and lemma 3.4 in the supplementary material, the r.h.s. tends to one. This proves the result for n, and the induction is complete.

Now, let us prove that $P(n \in \widehat{I}_1) \to 1_{n\in I_1}$ for all $n \in \mathbb{Z}$. We consider only $n > 0$, the proof being similar for negative values. When $n \in I_1$, the result follows directly from (8.2). When $n \notin I_1$, two cases can be distinguished. First, if $n-1 \in I_1$, $P(n \in \widehat{I}_1) \to 0$ by lemma 3.4 in the supplementary material. Second, if $n-1 \notin I_1$, let us consider $n_0 = \min\{n > 0 : n \notin I_1\}$. $n_0 - 1 \in I_1$, so that by lemma 3.4 again, $P(n_0 \in \widehat{I}_1) \to 0$. Moreover, by definition of $\widehat{I}_1$, $P(n \in \widehat{I}_1) \leq P(n_0 \in \widehat{I}_1)$. Thus $P(n \in \widehat{I}_1) \to 0$, which proves the result.

We now show that the estimators of $C'(y_n)$ and $F_\theta(\theta_n)$ are consistent. Because $\widehat{C'(y_n)} = t_1'(\widehat{y}_n)/\widehat{\theta}_n$ is a continuous function of $(\widehat{y}_n, \widehat{\theta}_n)$, (8.2) implies the convergence of $\widehat{C'(y_n)}$. Moreover, by the triangular inequality,

$$\begin{aligned} |\widehat{F_\theta(\theta_n)} - F_\theta(\theta_n)| &\leq |\widehat{F}_{y_1}(\widehat{y}_n) - F_{y_1}(\widehat{y}_n)| + |F_{y_1}(\widehat{y}_n) - F_{y_1}(y_n)| \\ &\leq \sup_{y\in\mathcal{Y}_1} |\widehat{F}_{y_1}(y) - F_{y_1}(y)| + |F_{y_1}(\widehat{y}_n) - F_{y_1}(y_n)|. \end{aligned}$$

The first term converges to zero by the Glivenko-Cantelli theorem. The second term also converges to zero by (8.2) and the continuity of F_{y_1}. Hence, $\widehat{F_\theta(\theta_n)}$ also converges.

2. Convergence of $\underline{\widehat{\Delta S_1}}$ and $\widehat{\overline{\Delta S_1}}$.

$t_1(\widehat{y}_n) - t_1(\widehat{y}_{n-1})$ and $\ln \widehat{\theta}_n - \ln \widehat{\theta}_{n-1}$ are consistent by what precedes and the continuous mapping theorem. Because $[\widehat{y}_{n-1}, \widehat{y}_n]$ is in a given compact set strictly included in $\mathcal{Y}_1$ with an arbitrary large probability, it follows from lemma 3.5 in the supplementary material that

$$\sup_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} |\widehat{q}_1(u) - q_1(u)| \xrightarrow{\mathbb{P}} 0.$$

This implies, by continuity of the infimum function with respect to the uniform norm,

$$\inf_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} \widehat{q}_1(u) - \inf_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} q_1(u) \xrightarrow{\mathbb{P}} 0 \tag{8.4}$$

Now, by continuity of $q_1(.)$, the mapping $(y, y') \mapsto \inf_{u\in[y',y]} q_1(u)$ is continuous. Hence,

$$\inf_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} q_1(u) \xrightarrow{\mathbb{P}} \inf_{u\in[y_{n-1},y_n]} q_1(u) \tag{8.5}$$

(8.4) and (8.5) implies the convergence of $\inf_{u\in[\widehat{y}_{n-1},\widehat{y}_n]} \widehat{q}_1(u)$, and consistency of $\widehat{\underline{\Delta S_1}}_n$ follows. The same reasoning applies to $\widehat{\overline{\Delta S_1}}_n$.

3. Consistency of the estimated bounds of C' and F_θ.

By what precedes, it suffices to prove the consistency of $\widehat{\underline{\theta}}_1(.), \widehat{\overline{\theta}}_1(.), \widehat{\underline{y}}_1(.)$ and $\widehat{\overline{y}}_1(.)$ on respectively $]\underline{y}_1, \overline{y}_1[$ and $]\underline{\theta}, \overline{\theta}[$. We focus on $\widehat{\underline{\theta}}_1(.)$ and $\widehat{\underline{y}}_1(.)$ here, the reasoning being similar for the upper bounds.

i. Let us begin with $\widehat{\underline{\theta}}_1(y) = \sup_{n\in\widehat{I}_1:\,\widehat{y}_n\geq y} \widehat{\theta}_n$ and suppose first that the set $\{n \in I_1 : y_n \geq y\}$ is not empty. We have that

$$\underline{\theta}_1(y) = \theta_{\underline{n}_1(y)},$$

where $\underline{n}_1(y) = \min\{n \in I_1 : y_n \geq y\}$. This stems from the facts that $(\theta_n)_{n\in I_1}$ is decreasing and $\inf\{n \in I_1 : y_n \geq y\} < -\infty$. If the last point were not true, $-\mathbb{N}$ would be included in I_1, and $(y_{-n})_{n\in\mathbb{N}}$, which is decreasing and bounded below by y, would converge to $y_{-\infty} \geq y$. This would imply $F_{y_1}(y_{-\infty}) = F_{y_2}(y_{-\infty})$, which is a contradiction since $F_{y_1} < F_{y_2}$ in the interior of $\mathcal{Y}_1$.

If $\underline{n}_1(y) > \inf I_1$, then, by definition of $\underline{n}_1(y)$ and because $y \notin \{y_n, n \in I_1\}$,

$$y_{\underline{n}_1(y)-1} < y < y_{\underline{n}_1(y)}.$$

Let us consider the event

$$E_1 = \left\{(\underline{n}_1(y) - 1, \underline{n}_1(y)) \in \widehat{I}_1^2, \widehat{y}_{\underline{n}_1(y)-1} < y < \widehat{y}_{\underline{n}_1(y)}\right\}.$$

By part 1 of the proof, $P(E_1) \to 1$. This proves the convergence in probability of $\widehat{\underline{\theta}}_1(y)$, since under E_1, $\widehat{\underline{\theta}}_1(y) = \underline{\theta}_1(y)$.

If $\underline{n}_1(y) = \inf I_1$, we consider similarly

$$E_1' = \left\{ \underline{n}_1(y) - 1 \notin \widehat{I}_1, \underline{n}_1(y) \in \widehat{I}_1, y < \widehat{y}_{\underline{n}_1(y)} \right) \}.$$

By part 1 of the proof once more, $P(E_1') \to 1$, which proves the convergence of $\widehat{\underline{\theta}}_1(y)$.

Suppose now that $\{n \in I_1 : y_n \geq y\} = \emptyset$. Then $\underline{\theta}_1(y) = 0$. Moreover, I_1 admits in this case a greatest element. Otherwise, indeed, we can show as previously that $(y_n)_{n\in\mathbb{N}}$ would tend to $y_\infty \leq y$, a contradiction. Let $\overline{n}_1$ denote this element, we have $y_{\overline{n}_1} < y$. By part 1 of the proof, the event $E_1'' = \{\overline{n}_1 \in \widehat{I}_1, \widehat{y}_{\overline{n}_1} < y\}$ has an arbitrarily large probability. The result follows because under E_1'', $\widehat{\underline{\theta}}_1(y) = 0$.

ii. We now turn to $\widehat{\underline{y}}_1(.)$. First, suppose that the set $\{n \in I_1 : \theta_n \geq \theta\}$ is not empty. Then,

$$\underline{y}_1(\theta) = y_{\overline{n}_1(\theta)},$$

where $\overline{n}_1(\theta) = \max\{n \in I_1 : \theta_n \geq \theta\}$. This stems from the facts that $(y_n)_{n\in I_1}$ is increasing and $\sup\{n \in I_1 : \theta_n \geq \theta\} < \infty$. If the last point were not true, $\theta \leq \theta_n = (\delta_1/\delta_2)^n \theta_0$ for all $n \geq 0$, so that $\theta \leq 0$, which is ruled out.

If $\overline{n}_1(\theta) < \sup I_1$, consider the event

$$E_2 = \left\{ \overline{n}_1(\theta), \overline{n}_1(\theta) + 1) \in \widehat{I}_1^2 \right\}.$$

Under E_2, we have, because $\widehat{\theta}_n = \theta_n$ for all $n \in I_1 \cap \widehat{I}_1$, $\widehat{\underline{y}}_1(\theta) = \widehat{y}_{\overline{n}_1(\theta)}$. Thus, for all $\varepsilon > 0$,

$$\begin{aligned} P(|\widehat{\underline{y}}_1(\theta) - \underline{y}_1(\theta)| > \varepsilon) &\leq P(E_2, |\widehat{y}_{\overline{n}_1(\theta)} - y_{\overline{n}_1(\theta)}| > \varepsilon) + P({}^cE_2) \\ &\leq P(\overline{n}_1(\theta) \in \widehat{I}_1, |\widehat{y}_{\overline{n}_1(\theta)} - y_{\overline{n}_1(\theta)}| > \varepsilon) + P({}^cE_2). \end{aligned}$$

By part 1 of the proof, $P({}^c E_2) \to 0$. Moreover, the first term tends to zero by (8.2), which proves the consistency of $\widehat{\underline{y}}_1(\theta)$.

It remains to consider the case when $\overline{n}_1(\theta) = \sup I_1$. The proof is identical except that we replace E_2 by

$$E_2' = \left\{ \overline{n}_1(\theta) \in \widehat{I}_1, \overline{n}_1(\theta) + 1 \notin \widehat{I}_1 \right\}.$$

Finally, if the set $\{n \in I_1 : \theta_n \geq \theta\}$ is empty, then $\underline{y}_1(\theta) = \underline{y}_1$. Because the sequence $((\delta_1/\delta_2)^{-n}\theta_0)_{n\in\mathbb{N}}$ tends to infinity, I_1 admits a smaller element $\underline{n}_1$. Let E_2'' denote the event $\{\underline{n}_1 \in \widehat{I}_1, \underline{n}_1 - 1 \notin \widehat{I}_1\}$. Under E_2'', we have $\widehat{\underline{y}}_1(\theta) = \widehat{\underline{y}}_1$. The result follows because E'' holds with an arbitrarily large probability and $\widehat{\underline{y}}_1$ converges in probability to $\underline{y}_1$. ■

Proof of Theorem 4.2

The proof is divided in four steps.

Step 1. Uniform convergence of the $(\widehat{y}_{k,l})_{(k,l)\in\widehat{I}_2}$.

Let $Y^\Delta = [\underline{y}-\Delta, \overline{y}+\Delta]$, we shall prove shortly that for all $\Delta > 0$ such that $\mathbb{Y}_1 \subset Y^{-\Delta} \subset Y^\Delta \subsetneq \mathcal{Y}_1$ and all $\varepsilon > 0$,

$$P\left(E_N(\varepsilon,\Delta)\right) = P\left(I_2^{-\Delta} \subset \widehat{I}_2 \subset I_2^\Delta,\ \max_{(k,l)\in\widehat{I}_2} |\widehat{y}_{k,l} - y_{k,l}| \leq \varepsilon\right) \to 1 \tag{8.6}$$

where we let for all Δ,

$$I_2^\Delta = \{(k,l) \in I_2 :\ |k| + |l| \leq m_N, y_{k,l} \in Y^\Delta\}.$$

To prove (8.6), let us introduce $M = 2 \vee \max_{j\neq k\in\{1,2,3\}^2} \sup_{x\in Y^\Delta} |H'_{kj}(x)|$, $\varepsilon_{m,N} = \varepsilon(M^m - 1)/(M^{m_N} - 1)$ and let $E'_N(\varepsilon,\Delta)$ be the event defined by

$$E'_N(\varepsilon,\Delta) = \left\{\forall j \neq k \in \{1,2,3\}^2,\ \sup_{x\in Y^\Delta} |\widehat{H}_{kj}(x) - H_{kj}(x)| \leq \varepsilon_{1,N}\right\} \tag{8.7}$$

Eventually, define the event

$$E_{m,N}(\varepsilon,\Delta) = \left\{I_{2,m}^{-\Delta} \subset \widehat{I}_{2,m} \subset I_{2,m}^\Delta,\ \max_{(k,l)\in\widehat{I}_{2,m}} |\widehat{y}_{k,l} - y_{k,l}| \leq \varepsilon_{m,N}\right\},$$

where $\widehat{I}_{2,m} = \{(k,l) \in \widehat{I}_2 :\ |k| + |l| = m\}$ and $I_{2,m}^\Delta, I_{2,m}^{-\Delta}$ are defined similarly.

Because $\varepsilon_{m,N} \leq \varepsilon$, we have

$$P(E_N(\varepsilon,\Delta) \geq P\left(\cup_{m=0}^{m_N} E_{m,N}(\varepsilon,\Delta)\right).$$

Moreover, assuming w.l.o.g. that $\varepsilon < \Delta$, we have, by lemma 3.6 in the supplementary material,

$$P\left(\cup_{m=0}^{m_N} E_{m,N}(\varepsilon,\Delta)\right) \geq P(E'_N(\varepsilon,\Delta)$$

Besides, because $m_N = o(\log(N))$, $1/\varepsilon_{m_N,N} = o(\sqrt{N})$. Thus, by lemma 3.3, $P(E'_N(\varepsilon,\Delta)) \to 1$. This proves (8.6).

Step 2. Uniform convergence of $\widehat{y}_1(.)$.

We just proved uniform convergence in probability of $\widehat{y}_1(\theta_{k,l})_{k,l\in\widehat{I}_2}$. Now we extend this result to the uniform convergence of $\widehat{y}_1(.)$ on $\Theta^{-\Delta} = \theta_1(Y^{-\Delta})$, for all $\Delta > 0$ such that

$\mathbb{Y}_1 \subset Y^{-\Delta} \subset Y^{\Delta} \subsetneq \mathcal{Y}_1$. Recall that for all θ, $\widehat{y}_1(\theta)$ is defined by linear interpolation, that is to say there exists $\lambda \in [0,1]$ such that

$$\widehat{y}_1(\theta) = \lambda \widehat{y}_1(\widehat{\underline{\theta}}(\theta)) + (1-\lambda)\widehat{y}_1(\widehat{\overline{\theta}}(\theta))$$

where

$$\begin{aligned}\widehat{\underline{\theta}}(\theta) &= \sup\left\{\theta_{k,l} : \theta_{k,l} \leq \theta, (k,l) \in \widehat{I}_2\right\} \text{ if this set is not empty} \\ &= \inf\{\theta_{k,l}, (k,l) \in \widehat{I}_2\} \qquad\qquad \text{otherwise}\end{aligned}$$

$\widehat{\overline{\theta}}(\theta)$ is defined similarly. Thus,

$$\begin{aligned}\sup_{\theta \in \Theta^{-\Delta}} |\widehat{y}_1(\theta) - y_1(\theta)| &\leq \sup_{\theta \in \Theta^{-\Delta}} \max\left\{|\widehat{y}_1(\widehat{\underline{\theta}}(\theta)) - y_1(\theta)|, |\widehat{y}_1(\widehat{\overline{\theta}}(\theta)) - y_1(\theta)|\right\} \\ &\leq \max\left\{\sup_{\theta \in \Theta^{-\Delta}} |\widehat{y}_1(\widehat{\underline{\theta}}(\theta)) - y_1(\theta)|, \sup_{\theta \in \Theta^{-\Delta}} |\widehat{y}_1(\widehat{\overline{\theta}}(\theta)) - y_1(\theta)|\right\}.\end{aligned}$$

Let us show that the first term in the maximum tends to zero in probability as N tends to infinity (the reasoning is similar for the second term). We have

$$\begin{aligned}\sup_{\theta \in \Theta^{-\Delta}} |\widehat{y}_1(\widehat{\underline{\theta}}(\theta)) - y_1(\theta)| &\leq \sup_{\theta \in \Theta^{-\Delta}} \left(\left|\widehat{y}_1(\widehat{\underline{\theta}}(\theta)) - y_1(\widehat{\underline{\theta}}(\theta))\right| + \left|y_1(\widehat{\underline{\theta}}(\theta)) - y_1(\theta)\right|\right) \\ &\leq \sup_{k,l \in \widehat{I}_2} |\widehat{y}_{k,l} - y_{k,l}| + \sup_{\theta \in \Theta^{-\Delta}} \left|y_1(\widehat{\underline{\theta}}(\theta)) - y_1(\theta)\right|\end{aligned}$$

The first term tends to zero in probability by step 1. Let us show that the second also tends to zero. Fix $\varepsilon > 0$. By uniform continuity of $y_1(.)$ on $\theta_1(Y^{-\Delta})$, there exists $\eta > 0$ such that

$$\sup_{(\theta,\theta') \in (\Theta^{-\Delta})^2 : |\theta - \theta'| < \eta} |y_1(\theta) - y_1(\theta')| < \varepsilon. \tag{8.8}$$

Thus, it remains to show that with probability approaching 1,

$$\sup_{\theta \in \Theta^{-\Delta}} |\widehat{\underline{\theta}}(\theta) - \theta| < \eta \tag{8.9}$$

To do so, let us introduce $0 < \Delta' < \Delta$ such that $\min \Theta^{-\Delta'} < \min \Theta^{\Delta} - \eta$. Using the density property established in lemma 3.1 in the supplementary material, and applying it to the compact interval $\Theta^{-\Delta'}$, and because $m_N \to \infty$, there exists N_0 and $(k,l) \in I_2^{-\Delta'}$ such that, for all $N \geq N_0$, $\theta_{k,l} < \theta$ and $|\theta_{k,l} - \theta| < \eta$.

Because $I_2^{-\Delta'} \subset \widehat{I}_2$ with probability approaching one (by step 1), we get, for all $\theta \in \Theta^{-\Delta}$,

$$0 < \theta - \widehat{\underline{\theta}}(\theta) \leq \theta - \theta_{k,l} < \eta.$$

This proves (8.9), and the uniform consistency of $\widehat{y}_1(.)$ follows.

Step 3. Uniform convergence of $\widehat{\theta}_1(.)$.
First, recall that for all $y \in Y^{-\Delta}$, there exists $\lambda \in [0,1]$ such that

$$\widehat{\theta}_1(y) = \lambda \widehat{\theta}_1(\underline{\widehat{y}}(y)) + (1-\lambda)\widehat{\theta}_1(\overline{\widehat{y}}(y))$$

where

$$\begin{aligned} \underline{\widehat{y}}(y) &= \sup\left\{\widehat{y}_{k,l} : \widehat{y}_{k,l} \leq y, (k,l) \in \widehat{I}_2\right\} \text{ if this set is not empty} \\ &= \inf\left\{\widehat{y}_{k,l} : (k,l) \in \widehat{I}_2\right\} \qquad \text{otherwise.} \end{aligned}$$

Then

$$\sup_{y\in Y^{-\Delta}} |\widehat{\theta}_1(y) - \theta_1(y)| \leq \max\left\{ \sup_{y\in Y^{-\Delta}} |\widehat{\theta}_1(\underline{\widehat{y}}(y)) - \theta_1(y)|, \sup_{y\in Y^{-\Delta}} |\widehat{\theta}_1(\overline{\widehat{y}}(y)) - \theta_1(y)| \right\} \quad (8.10)$$

Let us focus on the first term. Let $(\widehat{k}(y), \widehat{l}(y)) \in \widehat{I}_2$ be such that $\widehat{y}_{\widehat{k}(y),\widehat{l}(y)} = \underline{\widehat{y}}(y)$. Under the event $E_N(\varepsilon, \Delta)$, $(k,l) \in I^{\Delta}$, and

$$\begin{aligned} |\widehat{\theta}_1(\underline{\widehat{y}}(y)) - \theta_1(y)| &= |\widehat{\theta}_1(\widehat{y}_{\widehat{k}(y),\widehat{l}(y)}) - \theta_1(y)| \\ &= |\theta_1(y_{\widehat{k}(y),\widehat{l}(y)}) - \theta_1(y)| \\ &\leq |\theta_1(y_{\widehat{k}(y),\widehat{l}(y)}) - \theta_1(\widehat{y}_{\widehat{k}(y),\widehat{l}(y)})| + |\theta_1(\widehat{y}_{\widehat{k}(y),\widehat{l}(y)}) - \theta_1(y)| \quad (8.11) \end{aligned}$$

By step 1 and uniform continuity of $\theta_1(.)$, the first term tends to zero uniformly. Thus, it suffices to prove the uniform convergence to zero of the second term. By uniform continuity of $\theta_1(.)$ on Y^{Δ}, it suffices to show that for all $\eta > 0$, with probability approaching one,

$$\sup_{y\in Y^{-\Delta}} |\underline{\widehat{y}}(y) - y| < \varepsilon \quad (8.12)$$

Applying the density argument of lemma 3.1 again, it exists, with probability approaching one, $(k,l) \in \widehat{I}_2$ such that $y_{k,l} < y - \eta/4$ and $y_{k,l} > y - 3\eta/4$. By uniform convergence of the $(\widehat{y}_{k,l})_{(k,l)\in \widehat{I}_2}$, $|\widehat{y}_{k,l} - y_{k,l}| < \eta/4$ with probability approaching one. Hence,

$$0 < y - \widehat{y}_{\widehat{k}(y),\widehat{l}(y)} < y - \widehat{y}_{k,l} < \eta.$$

This proves (8.12), and the uniform consistency of $\widehat{\theta}_1(.)$ follows.

Step 4. Convergence of $\widehat{C}'(.)$, $\widehat{F}_\theta(.)$ and $\widehat{\Delta S}_1(.)$.

Firstly, convergence of $\widehat{C}'(.)$ and $\widehat{F}_\theta(.)$ follows using the same arguments as in Theorem 4.1. To establish consistency of $\widehat{\Delta S}_1(.)$, suppose w.l.o.g. that $y \in (y_0, \overline{y})$ and remark that

$$\begin{aligned} |\widehat{\Delta S}_1(y) - \Delta S_1(y)| &\leq \left|\widehat{q}_1(y)\ln\widehat{\theta}_1(y) - q_1(y)\ln\theta_1(y)\right| + |(\widehat{q}_1(y_0) - q_1(y_0))\ln\theta_0| \\ &\quad + \int_{y_0}^{y}\left|\widehat{q}_1'(u)\ln\widehat{\theta}_1(u) - q_1'(u)\ln\theta_1(u)\right| du. \end{aligned}$$

The first two terms tend to zero in probability by step 2, lemma 3.4 and the continuous mapping theorem. Let $\Delta > 0$ be such that $y \leq \overline{y} - \Delta$. The integral term I_N satisfies

$$\begin{aligned} I_N &\leq \int_{y_0}^{y} |\widehat{q}_1'(u)|\left|\ln\widehat{\theta}_1(u) - \ln\theta_1(u)\right| + |\ln\theta_1(u)|\,|\widehat{q}_1'(u) - q_1'(u)|\,du \\ &\leq \sup_{u\in Y^{-\Delta}}\left|\ln\widehat{\theta}_1(u) - \ln\theta_1(u)\right|\int_{y_0}^{y}|\widehat{q}_1'(u)du| + \sup_{u\in Y^{-\Delta}}|\widehat{q}_1'(u) - q_1'(u)|\int_{y_0}^{y}|\ln\theta_1(u)du| \end{aligned}$$

The first supremum in the r.h.s. tends to zero in probability by point 2 of the proof. The second supremum also tends to zero, by lemma 3.4. Finally, the first integral I_N' satisfies

$$I_N' \leq \int_{y_0}^{y}|q_1'(u)|\,du + (y - y_0)\sup_{u\in Y^{-\Delta}}|\widehat{q}_1'(u) - q_1'(u)|\,.$$

Thus, $I_N' = O_P(1)$. As a consequence, $I_N = o_P(1)$, which proves the convergence of $\widehat{\Delta S}_1(.)$. ■

References

Ausubel, L. M. (1999), Adverse selection in the credit card market. Mimeo.

Baron, D. P. and Myerson, R. B. (1982), 'Regulating a monopolist with unknown costs', *Econometrica* **50**, 911–930.

Chiappori, P. A. and Salanié, B. (2002), Testing contract theory: A survey of some recent work, *in* L. H. e. S. T. M. Dewatripont, ed., 'Advances in Economics and Econometrics', Vol. 1, New York: Cambridge University Press.

Chu and Sappington (2007), 'Simple cost-sharing contracts', *The American Economic Review* **97**, 419–428.

Crawford, G. S. and Shum, M. (2007), 'Monopoly quality degradation and regulation in cable television', *Journal of Law and Economics* **50**, 181–219.

D'Haultfoeuille, X. and Février, P. (2002), Persévérance, accessibilité et acceptation : une nouvelle modélisation de la non-réponse, *in* 'Actes des Journées de Méthodologie Statistique 2002', INSEE.

Ferrall, C. and Shearer, B. (1999), 'Incentives and transaction costs within the firm: Estimating an agency model using payroll records', *Review of Economic Studies* **99**, 309–338.

Freixas, X. and Laffont, J. J. (1990), Optimal banking contracts, *in* 'Essays in Honor of Edmond Malinvaud, vol. 2: Macroeconomics', MIT Press, pp. 33–61.

Gagnepain, P. and Ivaldi, M. (2002), 'Incentive regulatory policies: The case of public transit systems in france', *The RAND Journal of Economics* **33**, 605–629.

Gagnepain, P. and Ivaldi, M. (2007), Contract choice, incentives, and political capture in public transit. Mimeo.

Gence-Creux, C. (2000), Identification of non-additive structural functions, *in* 'Econometric Society World Congress 2000 Contributed Papers', New York: Cambridge University Press.

Guerre, E., Perrigne, I. and Vuong, Q. (2000), 'Optimal nonparametric estimation of first-price auctions', *Econometrica* **68**, 525–574.

Guerrre, E., Perrigne, I. and Vuong, Q. (2008), 'Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions', *Econometrica, forthcoming* .

Hahn, J., Todd, P. and van der Klauw, W. (2001), 'Identification and estimation of treatment effects with a regression-discontinuity design', *Econometrica* **69**, 201–209.

Huang, Y., Perrigne, I. and Vuong, Q. (2007), Nonlinear pricing in yellow pages. Mimeo.

Imbens, G. W. and Newey, W. K. (2009), 'Identification and estimation of triangular simultaneous equations models without additivity', *Econometrica, forthcoming* .

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

Ke, R. (2008), Identifying contract optimality non-parametrically with moral hazard: First order approach and statistical inference. Mimeo.

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

Laffont, J. J. and Tirole, J. (1986), 'Using cost observation to regulate firms', *The Journal of Political Economy* **94**, 614–641.

Lavergne, P. and Thomas, A. (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.

Le Lan, R. (2008), Enquêtes ménages de l'insee : vers la fin de la baisse des taux de réponse? Insee, internal study.

Leslie, P. (2004), 'Price discrimination in broadway theater', *The RAND Journal of Economics* **35**, 520–541.

Lewis, T. and Sapppington, D. (1995), 'Optimal capital structure in agency relationships', *The RAND Journal of Economics* **26**, 343–361.

Manning, W., Newhouse, J., Duan, N., Keeler, E. and Leibowitz, A. (1987), 'Health insurance and the demand for medical care: Evidence from a randomized experiment', *American Economic Review* **77**, 251–277.

Maskin, E. and Riley, J. (1984), 'Monopoly with incomplete information', *The RAND Journal of Economics* **15**, 171–196.

Miravete, E. J. (2002), 'Estimating demand for local telephone service with asymmetric information and optional calling plans', *Review of Economic Studies* **69**, 943–971.

Miravete, E. J. (2007), The limited gains from complex tariffs. Mimeo.

Miravete, E. J. and Roller, L.-H. (2005), 'Estimating markups under nonlinear pricing competition', *Journal of the European Economic Association* **2**, 526–535.

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

Myerson, R. B. (1979), 'Incentive compatibility and the bargaining problem', *Econometrica* **47**, 61–73.

Myerson, R. B. (1981), 'Optimal auction design', *Mathematics of Operation Research* **6**, 58–73.

Neeman, Z. (2003), 'The effectiveness of english auctions', *Games and Economic Behavior* **43**, 214–238.

Paarsch, L. (1992), 'Deciding between the common and private value paradigms in empirical models of auctions', *Journal of Econometrics* **51**, 191–215.

Paarsch, L. and Shearer, B. (2000), 'Piece rates, fixed wages and incentive effects: Statistical evidence from payroll records', *International Economic Review* **41**, 59–92.

Perrigne, I. (2002), Incentive regulatory contracts in public transportation: An empirical study. Working Paper, Pennsylvania State University.

Perrigne, I. and Vuong, Q. (2004), Nonparametric identification of incentive regulation models. Mimeo.

Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Literature* **37**, 7–63.

Rogerson, W. P. (2003), 'Simple menus of contracts in cost-based procurement and regulation simple menus of contracts in cost-based procurement and regulation', *The American Economic Review* **93**, 919–926.

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.