The provision of wage incentives: A structural estimation using contracts variation

Xavier D'Haultfœuille
CREST

Philippe Février
CREST

We address empirically the issues of the optimality of simple linear compensation contracts and the importance of asymmetries between firms and workers. For that purpose, we consider contracts between the French National Institute of Statistics and Economics (Insee) and the interviewers it hired to conduct its surveys in 2001, 2002, and 2003. To derive our results, we exploit an exogenous change in the contract structure in 2003, the piece rate increasing from 20.2 to 22.9 euros. We argue that such a change is crucial for a structural analysis. It allows us, in particular, to identify and recover nonparametrically some information on the cost function of the interviewers and on the distribution of their types. This information is used to select correctly our parametric restrictions. Our results indicate that the loss of using such simple contracts instead of the optimal ones is no more than 16%, which might explain why linear contracts are so popular. We also find moderate costs of asymmetric information in our data, the loss being around 22% of what Insee could achieve under complete information.

Keywords. Incentives, asymmetric information, optimal contracts, nonparametric identification.

JEL classification. C14, D82, D86.

1. Introduction

Over the past three decades, extensive attention has been devoted to asymmetries of information and their consequences in economics. These asymmetries play, in particular, a fundamental role in the economics of the firms (see Prendergast (1999) for a survey). Firms have to provide the right incentives to their workers, and design appropriate compensation plans, even when restricting to simple contracts such as piece rate, commissions at quota, or lump-sum bonuses. Indeed, a growing empirical literature shows that overall, incentives substantially increase workers' productivity (see, e.g., Lazear (2000)

or Paarsch and Shearer (2000)), and that the form of the payment scheme matters (Ferrall and Shearer (1999), Copeland and Monnet (2009), Chung, Steenburgh, and Sudhir (2014)). Our paper adds to this empirical personnel literature by quantifying the loss of using simple linear compensation contracts instead of nonlinear, optimal ones, and the importance of asymmetries between firms and workers.

We use for this purpose contract data between the French National Institute of Economics and Statistics (Insee) and its interviewers. Insee is a public institute that conducts each year between twelve and twenty household surveys on different topics such as labor force, consumption or health. It hires interviewers to contact the households and conduct the corresponding interviews. We have data on three successive surveys on household living conditions ("enquête Permanente sur les Conditions de Vie des Ménages," PCV hereafter) that took place in October 2001, 2002, and 2003. For each survey and all interviewers, we observe their average response rates, defined as the ratio of the number of respondents to the number of households each interviewer has to interview.

These response rates vary with the effort the interviewers make to contact the households and to persuade them to accept the interview. Response rates also differ from one interviewer to another because of the heterogeneity in interviewers' cost of effort, and differences between the geographical areas attributed to them. These unobserved effort and heterogeneity are the reasons why Insee faces an asymmetric information problem. To give incentives to its interviewers, Insee then uses a simple compensation scheme. Interviewers receive a basic wage (around 4.7 euros in the three surveys), which does not depend on whether the interview is achieved or not, plus a bonus for each interview they conduct. The key point of the paper is to exploit the fact that the bonus changed in 2003, increasing from 20.2 euros in 2001 and 2002 to 22.9 euros in 2003. Moreover, we have reasons to believe that this increase was not due to a change in the cost of interviewers.

To investigate the efficiency of simple linear compensation contracts and the importance of asymmetries between Insee and its interviewers, we rely on a structural principal-agent model that incorporates both adverse selection and moral hazard. We show that the cost function and the distribution of the interviewers' types are partially identified nonparametrically using the exogenous change in contracts. An important feature of this result is that the information on the functions of interest are recovered using the interviewers' program solely. This is convenient because it is very likely that Insee does not implement the optimal contracts, but only optimizes over linear ones. More generally, aiming at testing the optimality of the principal precludes any identification method relying precisely on this optimality. Importantly, also, our identification result is robust to the presence of selection effects, namely whether or not the new compensation scheme has attracted better interviewers.

If the identification argument developed for the moral hazard part is specific, our result on the adverse selection part could actually apply to many adverse selection models, including regulatory contracts, and nonlinear and price discrimination models.[1] All of these models share a common underlying structure for which our procedure

---

[1]For an incomplete list of empirical papers in these fields, see Ivaldi and Martimort (1994), Wolak (1994), Gagnepain and Ivaldi (2002), Miravete (2002), Leslie (2004), Miravete and Roller (2005), Lavergne

is well adapted and can be useful to study their nonparametric identification. Though the models somewhat differ, our identification result is therefore connected with those of Perrigne and Vuong (2011b), Aryal, Perrigne, and Vuong (2016), Luo, Perrigne, and Vuong (2018), and Aryal and Gabrielli (2018) on regulation, insurance models, unidimensional and multidimensional nonlinear pricing, respectively. An important difference with these papers is that we neither rely on the knowledge of the principal's objective function, nor on the optimality of observed contracts. On the other hand, the identification of the cost function and the distribution of the interviewers' types relies on exogenous variation in contracts, and is only partial with one exogenous change.

Our identification argument is also related to the identification of first-price auctions models with risk-adverse bidders, using exogenous variations in the number of bidders (Guerre, Perrigne, and Vuong (2009)). Interestingly also, we show that our problem boils down to the identification of nonparametric transformation models or, equivalently in duration models, generalized accelerated failure time model, with discrete regressors. This question has been studied by Abbring and Ridder (2015), but under some large support conditions and regularity conditions at the boundary of this support. We show that without such conditions, the model is still partially identified.

Beyond identification, we also develop a nonparametric estimation procedure using our identification method. We estimate nonparametrically bounds on the cost function and the distribution of interviewers' type. In a second step, we introduce parametric specifications in line with the nonparametric estimates of the interviewers' cost function and distribution of types. As the model is not point identified nonparametrically, such restrictions are necessary to estimate the policy effects we are interested in. However, contrary to most papers in the personnel literature, which adopt directly a parametric framework, our specifications are driven by the nonparametric analysis.

Studying Insee and its interviewers, our method allows us, first, to conclude that the loss of using a simple contract instead of an optimal one is rather small, around 16%. Even if the theoretical literature concludes that optimal contracts are in general nonlinear (see Laffont and Martimort (2002), for a survey),[2] simple compensation schemes such as piece rates and bonuses are usually thought of as the best compromise between efficiency and ease of implementation (Raju and Srinivasan (1996)). Our result supports this claim and may explain why simple contracts are so popular and widely used by firms. This idea is also in line with the theoretical findings of Wilson (1993, Section 6.4), Rogerson (2003), and Chu and Sappington (2007), who show that simple tariffs can secure more than 70% of the maximal surplus. Firms can adopt simple compensation systems and still give the right incentives to workers. Little empirical work has however tried to estimate the loss associated with the use of simple compensation scheme and the empirical personnel literature mentioned previously usually abstracts from these issues. An exception is Miravete (2007), who reports a loss of only 3%. Ferrall and Shearer (1999), on the other hand, concluded that simple nonlinear compensation plans lead to substantial inefficiencies.

and Thomas (2005), Crawford and Shum (2007), Perrigne and Vuong (2011a), Miravete (2007), Gagnepain, Ivaldi, and Martimort (2013), Lim and Yurukoglu (2018), and Kang and Silveira (2018).

[2]An exception is the result of Holmstrom and Milgrom (1987).

Our method also allows us to recover what Insee's surplus would have been under complete information. Independently of the issue of contracts' optimality, asymmetries create inefficiencies because of the informational rent captured by the agents. Measuring this rent is therefore important for the firm. This question is central in the insurance literature (see Chiappori and Salanié (2003), for a survey), or in the auction literature (see Perrigne and Vuong (1999), for a survey). On the contrary, few empirical works have focused on quantifying the magnitude of such asymmetries between firms and workers in the personnel literature. We find moderate cost of asymmetric information, the estimated expected surplus under incomplete information being 78% of the full information surplus. This loss (22%) is in particular smaller than the one reported by Ferrall and Shearer (1999) who found an efficiency loss of 33%. Overall, in our data, the surplus under asymmetric information and with a simple linear compensation plan is 66% of what it could be under complete information. The main part of this loss (65%) is due to incomplete information whereas the last 35% are associated with the simple payment scheme.

The paper is organized as follows. Section 2 presents institutional details and the data at our disposal. In Section 3, we focus on the interviewers' behavior. We develop a simple theoretical model and show that it is partially identified thanks to the exogenous change in the contract. We then propose estimators for the corresponding bounds and show their consistency. Finally, we estimate these bounds on the data. Section 4 focuses on the policy analysis. We show how the information on interviewers can be used to recover counterfactual parameters. We then study the optimality of the linear contracts used by Insee and the importance of asymmetries in this context. Section 5 concludes.

2. Institutional details and data description

The French National Institute of Economics and Statistics (Insee) conducts each year between twelve and twenty household surveys on different topics such as labor force, consumption, or health. For that purpose, Insee used to draw, until 2009 and approximately every 10 years, a large sample of housings from the exhaustive census database. This sample consisted of geographical areas called primary units. All survey samples were then drawn from these primary units. To conduct the interviews, Insee hired interviewers who live close to the primary units, in order to limit their traveling costs.

Interviewers' work is similar for almost all surveys. First, Insee gives them a list of sampled households to interview in their designated area, as well as some characteristics of the housings and households, as described in the census database. Interviewers then have to locate precisely the housings of their sample in order, for instance, to identify unoccupied or destroyed housings. After that, they try to contact the households. This stage is the main part of their job and usually takes several days. Usually, interviewers have to go to the housings several times and leave phone messages before coming in contact with the household. Finally, once contacted, interviewers have to convince the households to accept the survey. In theory, it is usually mandatory to participate to a survey by Insee. In practice, during the period we consider hereafter, more than 90% of

households accepted to participate, once they had been contacted.[3] In a typical household survey, it takes around one hour to go through all the questions. In compensation, interviewers were paid in a similar way for all household surveys until 2013. They received a basic wage for each household they have to interview, plus a bonus for each interview they achieved. They were also reimbursed for all their expenses, such as the travel costs or the meals they have to take during their work.

We have data on three successive surveys on household living conditions ("enquête Permanente sur les Conditions de Vie des Ménages", PCV hereafter), which took place in October 2001, 2002, and 2003.[4] Each survey comprises a fixed part, which is identical for each edition (representing more than half of the questions), and a complementary part, which changes every year. In 2001, 2002, and 2003, the focus of the survey was put respectively on the use of new technologies, participation in associations and education practices in the family. For each survey, our dataset consists of the list of all housings in the survey sample, excluding secondary, unoccupied and destroyed housings. For each housing, we observe some of its characteristics in the 1999 census, namely the number of rooms, the household size, and the age of the reference person. We also observe the identification number of the interviewer in charge of interviewing the corresponding household, and a dummy indicating whether the interview was conducted or not. Table 1 summarizes the main information about the three surveys, on the whole sample of households. There were between 379 and 478 interviewers in each survey. On average, each interviewer was assigned around 16 households in 2001 and 2002, and 28 in 2003.

The 2001 and 2002 surveys display very similar patterns. In particular, their average response rates, defined as the ratio of the number of respondents to the number of housings, are not significantly different at the 5% level (78.5 and 77.7%, resp.). Their distribution functions are also very close (see Figure 1), with a p-value of the two-sided Kolmogorov–Smirnov test equal to 0.87. On the other hand, the average response rate is significantly higher in 2003 (80.7%), and the distribution function of the 2003 survey stochastically dominates the one of 2001–2002[5] (see Figure 1), with a p-value of the one-sided Kolmogorov–Smirnov test equal to 0.003. We also note that the distribution functions displayed in Figure 1 exhibit several jumps, especially at 0.5, 0.67, and 1. These

TABLE 1. Descriptive statistics on the full sample.

Year	Number of Interviewers	Number of Households	Average Response Rate
2001	379	17.3	78.5%
2002	478	15.4	77.7%
2003	453	28.0	80.7%

[3]Insee never fines households that do not participate, but interviewers can use the argument that the survey is mandatory to convince households to participate.

[4]We also have some limited information on interviewers that we use at the end of our analysis; see Appendix A for details on these data.

[5]The average response rate on 2001–2002 is defined as the ratio between the total number of interviews and the total number of households, where the 2001 and 2002 data are pooled.
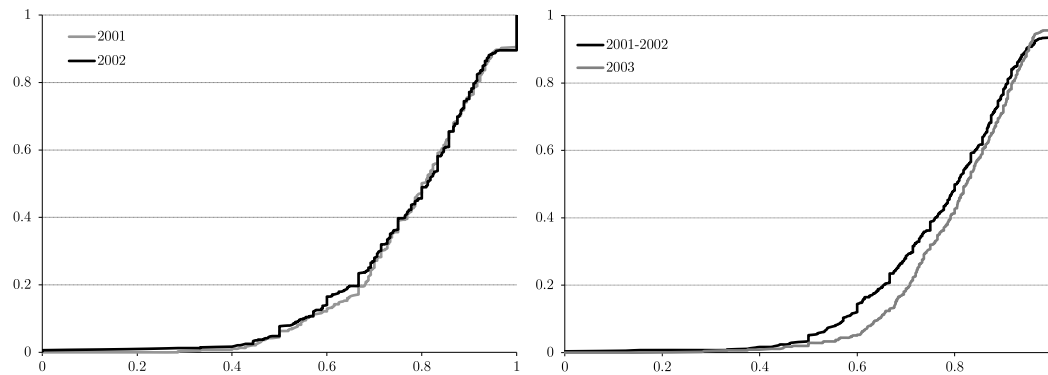
FIGURE 1. Distribution functions of the response rates on all interviewers, for all households.

jumps are due to the fact that the response rates are ratios of two integers, and the number of households to interview is rather small.[6]

There are two main differences between the 2003 and the other two surveys. The first one is related to its sampling design, and the second to its payment scheme. As previously mentioned, the PCV surveys are drawn from primary units. This was the case for the three surveys we consider. However, the sample was approximately twice as large in 2003 as in 2001 and 2002. Besides, because the 2003 survey focused on families, housings in which a family lived at the time of the census were overrepresented in 2003. As a result of this overrepresentation, housings in which a family lived at the time of the census represent 54.5% of the housings in 2003, as opposed to 44.4% and 48.3% in 2001 and 2002. Because families are on average easier to contact than, for instance, single persons, this difference may partly explain why response rates were higher in 2003. To control for this sampling effect and make comparisons possible for the three surveys, we restrict hereafter our attention to such housings occupied by families. These were the only differences in the survey designs of the three surveys. In particular, the corresponding subsamples of families were drawn similarly.

Table 2 shows that, as expected, the average response rates for families are higher than in the general population (resp., 79.0%, 79.8%, and 83.1% versus 78.5%, 77.7%,

TABLE 2. Descriptive statistics on the subsample of families.

Year	Number of Interviewers	Number of Families	Average Response Rate	Payment per Household		Average Income		
				Basic	Bonus	Basic	Bonus	Total
2001	377	8.35	79.0%	4.7	20.3	39.3	135.0	174.3
2002	471	6.85	79.8%	4.7	20.2	32.2	111.9	144.1
2003	453	15.24	83.1%	4.6	22.9	70.1	289.7	359.8

---

[6]Because of this small numbers of households, it is logical, from a pure statistical point of view, to observe more jumps at 0.5 or 0.67 as more integers can be divided by 2 or 3.

and 80.7%). When comparing the three surveys on families only, we find however the same pattern as in Table 1. The difference between the 2001 and 2002 surveys is not significant (79.0% and 79.8%, resp.), whereas interviewers achieve significantly higher response rates in 2003 (83.1%).

There is also a second difference in the three surveys, namely their payment schemes. Whereas the basic wage is nearly constant the 3 years, at a low level (4.7 euros in 2001, 4.6 euros in 2002 and 2003),[7] the bonus for achieving an interview with a family was 22.9 euros in 2003, compared to 20.3 and 20.2 euros in 2001 and 2002. We use this modification afterwards to identify the principal-agent model that we consider in the following section.

3. The interviewer's model

We first model the interviewers' decision, in particular to recover their utility function. We use this utility function in the next section to quantify the loss due to linear contracts and asymmetric information.

3.1 *The interviewers' program*

We suppose that interviewers decide on the effort they spend to try to contact each household. Instead of modeling effort, we model directly the probability of contact that each interviewer fixes for each household. These households are heterogeneous and may be easy or difficult to contact, depending on their characteristics. Single persons living in urban areas are difficult to contact, for instance, because they spend relatively little time at home, and digital locks make a direct contact more difficult to establish. Interviewers do not face such barriers in the countryside, and families are on average more at home. Once we restrict our attention to an interviewer's area and to the housings in which a family was living in 1999, however, households appear to be almost homogeneous ex ante. To support this claim, we regress the response rates of interviewers on the mean of the 1999 census characteristics (household size, number of rooms, and age of the reference person), controlling for interviewers and years fixed effects. While household size has a positive and significant effect when considering the whole sample, this effect disappears when restricting to the sample of families. None of the other census variables are significantly different from zero. As each interviewer works in a small and specific geographic area, this result does not really come as a surprise. In each restricted area, housings in which a family was living are, ex ante, quite similar and homogeneous for the interviewers.

Because families are homogeneous in terms of contact ease, we suppose that interviewers treat them similarly and take the same decision for all of them. An interviewer thus decides, for each household, with which probability y he wants to survey it, and produces his effort accordingly. As detailed below, this probability is not equal to the actual response rate because of randomness in interviewers' work. The expectation of the cost to reach a probability y is supposed to depend on the survey x, the number n

---

[7]All figures are in 2002 euros.

of households to interview but also the area and the interviewer herself. We summarize by θ the heterogeneity term in cost related to the interviewers and their areas. For simplicity, we refer subsequently to interviewers' type, but one should keep in mind this dual aspect of θ. At the end, we denote by $C(n, x, y, \theta)$ the expected cost of reaching a probability y in survey x, for an interviewer of type θ with n interviews to conduct.

To give the interviewers incentives to achieve high response rates, Insee provides them with a bonus if they realize the interview. Let $\delta(x)$ and $w(x)$ denote respectively the bonus and basic wage chosen by Insee for survey x. In this case, the interviewer receives $w(x) + \delta(x)$ when the interview is achieved and $w(x)$ otherwise. Hence, if the interviewer with n households to interview implements a probability y of conducting the survey for each household in his sample, he obtains on average a total wage of $n(\delta(x)y + w(x))$. We suppose hereafter that interviewers are risk-neutral and have a quasi-linear utility function. In this case, an interviewer of type θ chooses a probability $y(n, x, \theta)$ satisfying

$$y(n, x, \theta) \in \arg\max_y n[w(x) + \delta(x)y] - C(n, x, y, \theta). \tag{3.1}$$

We denote by $Y = y(N, X, \theta)$ the actual probability chosen by the interviewer.[8] Y is not observed by the principal Insee (nor the econometrician), which is the source of moral hazard here.[9] Instead, it only observes the number of interviews R the interviewer eventually does in survey X. Our first assumption relates this observable output R with Y.

ASSUMPTION 1 (Independence in households reactions). $R|N, X, Y \sim \text{Binomial}(N, Y)$.

We thus suppose that each household reacts independently from each other. Independence between households seems very likely here, as the households to interview are not neighbors in general, contrary to what happens in labor force surveys for instance.

Next, we impose a separability and regularity conditions on the cost of interviewers. To cope with potential selection effects, we introduce here S, the dummy of being a "stayer." Specifically, $S = 1$ if the interviewer participates to the 2003 and either the 2001 or 2002 surveys, $S = 0$ otherwise. Besides, for any random variables U and V, we let F_U (resp., $F_{U|V}$) denote the cumulative distribution function of U (resp., of U conditional on V). Finally, with a slight abuse of notation, we denote by $[a, b]$ closed intervals of the real line, even if b is possibly infinite.

ASSUMPTION 2 (Cost separability and continuous distribution of types). $C(n, x, y, \theta) = \theta C(n, x, y)$, where $C(n, x, \cdot)$ is twice continuously differentiable with $\partial^2 C/\partial y^2 > 0$ for all $y \in (0, 1)$. Moreover, $F_{\theta|X,S}$ has support $[\underline{\theta}, \overline{\theta}]$ with $0 \leq \underline{\theta} \leq \overline{\theta} \leq +\infty$ and it is continuously differentiable with density $f_{\theta|X,S}$.

---

[8]As usually, capital Latin letters correspond to random variables, while their lowercase counterpart are realizations of these variables.

[9]Even if we assume that interviewers are risk-neutral, moral hazard affects the design of contracts, because it means that Insee can only design contracts based on N and R, rather than on N and Y. We come back to this issue in Section 4.2.

Assumption 2 imposes in particular that the cost function is separable between the individual-area effect θ_i and a common cost of choosing the probability y. This cost separability assumption reduces the dimensionality of the problem and is necessary to identify the model (see D'Haultfœuille and Février (2010), for a discussion on this assumption). Such an 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), or Lavergne and Thomas (2005)).

Finally, we impose the following condition on the subsample size.

ASSUMPTION 3 (No effect and randomness of subsample size). *For all $n \in \text{Supp}(N|X = x)$, $C(n, x, y) = nC(x, y)$. Moreover, $N \perp\!\!\!\perp \theta | X, S$.*

The first condition implies that there are neither economies nor diseconomies of scale here. As the sample size grows, the interviewer may decrease the cost of each contact attempt by trying to meet several households located next to each other at the same time. But it may also be more difficult to find a convenient moment to get in touch with a household if many interviews have to be done during the day. The condition $C(n, x, y) = nC(x, y)$ is then reasonable if both aspects are of second order, or offset each other.[10] Assumption 3 also imposes that conditional on the survey and interviewers' status S, the number of interviews given to an interviewer is independent of his type.[11] This condition seems plausible because the sample is drawn at the national level and the general rule is that each interviewer receives the sample that corresponds to his geographic area.

It could still be the case, however, that when possible, Insee allocates some of the households to its best interviewers. Also, economies or diseconomies of scale may violate the condition $C(n, x, y) = nC(x, y)$. But importantly, Assumption 3 has testable implications. We develop in Section 3.5 and Appendix B.1 several tests of this condition, which all lend support to its validity.

The previous assumptions imply that the probability chosen by the interviewer does not depend on n and is defined by the first-order condition of (3.1). In other words, $y(x, \theta)$ satisfies

$$\delta(x) = \theta \frac{\partial C}{\partial y}\big(x, y(x, \theta)\big). \tag{3.2}$$

Moreover, differentiating this condition shows that $\theta \mapsto y(x, \theta)$ is strictly decreasing. We denote by $\theta(x, \cdot)$ its inverse function.

3.2 *Nonparametric identification*

We now turn to the empirical content of the interviewer's model. An observation is a vector (N, R, X, S).[12] For identification, we consider an ideal framework where the number

---

[10]In Appendix B.2, we discuss identification of marginal costs with only $C(n, x, y) = f(n)C(x, y)$ for some unknown $f(\cdot)$.

[11]On the other hand, and because of potential selection effects, we do not impose that the distribution of θ is the same for the two surveys, nor between stayers and "movers."

[12]Implicitly, we are therefore not exploiting the identity of interviewers, other than through the information of being a stayer or not. An advantage of this is that we do not restrict θ to be the same from one survey

of observations is infinite. Hence, we suppose that the conditional distribution function $F_{N,R|X,S}$ is known. The question is whether the marginal cost functions C' and the distribution of types $F_{\theta|X,S}$ can be recovered from these functions and the model. We decompose identification into two steps.

3.2.1 *Identification of the conditional distribution of Y*   Our first result is that the conditional distribution of Y is identified from the conditional distribution of (N, R) under our assumptions above and a large support condition.

ASSUMPTION 4 (Large support on N). $\sup \text{Supp}(N|X, S) = +\infty$ *with probability one.*

THEOREM 3.1. *If Assumptions* 1–4 *hold*, $F_{Y|X,S}$ *is identified from the distribution of* $(N, R)|X, S$.

The proof of Theorem 3.1, like all other proofs, is displayed in Appendix C. This result means that from an econometric point of view, we can ignore subsequently moral hazard, and makes as if the probabilities chosen by interviewers were observed. The proof of this result can be summarized as follows. Conditional on $N = n$ and $X = x$, R follows a binomial mixture model (see, e.g., Wood (1999) and for a recent application, D'Haultfœuille and Rathelot (2017)). In such models, the first n conditional moments of Y are identified. By independence between Y and N and because n can be chosen arbitrarily large, this implies that all conditional moments of Y are identified. The result follows because the distribution of a bounded variable is uniquely determined by its moments (see, e.g., Gut (2005)).

Assumption 4 may seem restrictive, because in practice the number of households given to an interviewer cannot exceed a certain threshold. Note however that If $\overline{N}_{x,s} \equiv \sup \text{Supp}(N|X = x, S = s) < +\infty$, we still identify the first $\overline{N}_{x,s}$ moments of $F_{Y|X,S}$ (in our data, $\overline{N}_{x,s} \geq 50$). We discuss in more detail in Appendix B.3 the consequence of a bounded support on N. In particular, we show therein that the approach taken by D'Haultfœuille and Rathelot (2017), who study the partial identification of functionals of $F_{Y|X,S}$ in the same binomial mixture model, leads to very similar estimators of $F_{Y|X,S}$. Hence, this issue does not seem to be of first order here.

3.2.2 *Identification of the marginal cost and distribution of interviewers' type*   We now turn to the identification of C' and $F_{\theta|X,S}$. We intensively use hereafter the assumption that the bonus changes exogenously.

ASSUMPTION 5 (Exogenous change in the contracts). $C(x, y) = C(y)$ *and* $X \perp\!\!\!\perp \theta|S = 1$.

We thus suppose that the change in the bonus observed in 2003 is not related to an unobserved change in interviewers' working condition. A first reason why Assumption 5 could fail is that households were more difficult to contact in surveys in 2003. This would explain why Insee increased its bonus this particular year. However, we would observe in

---

to another, for a given interviewer. We only assume in Assumption 5 below that the cdf of θ for stayers remains the same between the different surveys.

this case a similar increase in other household surveys, which is not the case. The compensation schemes of the two other regular surveys (namely the labor force survey and the survey on rents and service charges) that also took place in October 2003 were not modified.[13] A specific trend in the difficulties to contact households in the PCV surveys seems also unlikely. First, such a trend would be at odds with the fact that the bonus remained unchanged in 2002. Second, and as mentioned before, the distributions of the response rates observed in 2001 and 2002 are very close and not statistically different.

Hence, any change in the cost of interviewers that may have occurred in 2003 should be specific to this particular PCV survey. But as explained above, once we restrict to housings occupied by families, the October 2003 PCV survey was drawn with the same sampling design. Hence, the distribution of households' characteristics are the same each year, up to the sampling randomness. We can actually test for this, by checking whether the average characteristics of the housings attributed to interviewers differ systematically between 2001–2002 and 2003. These tests, detailed in Section 3.5 below, fully support this claim. Besides, the survey was conducted during the same period during the same period, namely the whole month of October. So we do not expect the households to be more or less absent from their house, because of a long vacation for instance, in 2003. Also, the 2003 survey had identical rules for the fieldwork than its predecessors.

Another explanation of a potential change in the cost in 2003 would be that the acceptance rate dropped because of the topic of the survey. However, the acceptance rates are rather constant over time, between 90 and 93% in the PCV surveys between 1999 and 2006 (Le Lan (2009)). These rates are very high as these surveys are usually mandatory and done by a public institute. Moreover, they do not vary much over time because the willingness to participate in a survey is mainly related to the time households have at their disposal (Le Lan (2009)). Hence, 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 questions.

Finally, even if the distribution of households to interview remained the same in 2003, Insee may have modified the assignment between interviewers and areas in 2003, resulting in a change in the distribution of θ. This would however be very much at odds with Insee's whole recruitment process. As explained above, Insee hires interviewers that live close to the primary units, in order to limit their traveling costs. There may also be a benefit of knowing well the area where one has to conduct interviews. That said, we can perform some suggestive tests, by looking whether experienced interviewers, say, had a different distribution of housing characteristics in 2003 than in 2001–2002. The tests, detailed in Section 3.5 below, strongly support this hypothesis.

For all of these reasons, we believe that the increase of the bonus in 2003 is not related to a change in the cost function of interviewers. Rather, we believe that it is due to

---

[13]The response rates of other Insee's surveys may be difficult to compare directly from one year to another because they may change in several dimensions. But if we consider the Labor Force Survey, which like the PCV survey, remains identical in 2002 and 2003, its response rate is very stable between the 2 years: 76.0% versus 75.5% for the third quarter, 81.3% versus 81.8% for the fourth quarter.

an increase in the "social value" of the information in the 2003 survey.[14] That the cost function did not change does not mean that incentive effects entirely explain the pattern observed in Figure 1. As mentioned before, the 2003 compensation scheme may have attracted more efficient interviewers. Such a selection effect is compatible with Assumption 5, as this condition remains silent on the selection process of interviewers.

   Assumption 5 and the fact that the 2001 and 2002 surveys have identical bonuses implies that interviewers choose the same probability for both years. Together with Assumptions 1 and 3, this implies that we can gather together the 2001 and 2002 surveys, making as if it was a single survey, since the total number of respondents among the two surveys will still be binomial. $X = 1$ refers hereafter to the 2001–2002 surveys, while $X = 2$ corresponds to the 2003 survey.

   Before studying identification of C' and $F_{\theta|X,S}$, note that a normalization is necessary since for any $\alpha > 0$, we can replace (θ, C') by $(\alpha\theta, C'/\alpha)$ and leave the model unchanged. Hence, for any $c_0 > 0$, we can choose $y_0 \in \mathcal{Y}_1 \equiv \mathrm{Supp}(Y|X = 1)$ so that $C'(y_0) = c_0$. Now, the idea for identification is to use the first-order condition (3.2) together with the exogenous changes in contracts. We also focus on stayers ($S = 1$), for whom there is no selection issue. First, for any $\tau \in (0, 1)$,

$$
\begin{aligned}
C'\big(F^{-1}_{Y|X=2,S=1}(\tau)\big) &= F^{-1}_{C'(Y)|X=2,S=1}(\tau) \\
&= \delta(2)F^{-1}_{1/\theta|X=2,S=1}(\tau) \\
&= \frac{\delta(2)}{\delta(1)}\delta(1)F^{-1}_{1/\theta|X=1,S=1}(\tau) \\
&= \frac{\delta(2)}{\delta(1)}C'\big(F^{-1}_{Y|X=1,S=1}(\tau)\big).
\end{aligned}
$$

The first equality holds by strict monotonicity of C', the second by the first-order condition (3.2), the third by $X \perp\!\!\!\perp \theta|S = 1$, and the fourth by applying the same reasoning with $X = 1$ instead of $X = 2$. Let $H(y) = F^{-1}_{Y|X=2,S=1} \circ F_{Y|X=1,S=1}(y)$ denote the quantile-quantile transform between the two conditional distributions of stayers. Remark that by Assumptions 2–3, $\mathcal{Y}_1 = \mathrm{Supp}(Y|X = 1, S = 1)$, and so H is defined on $\mathcal{Y}_1$. Moreover, for all $y \in \mathcal{Y}_1$,

$$
C'\big(H(y)\big) = \frac{\delta(2)}{\delta(1)}C'(y). \tag{3.3}
$$

By Theorem 3.1, H is identified under Assumptions 1–4. Then (3.3) shows that if C' is identified at $y \in \mathcal{Y}_1$, it is also identified at $H(y)$. We use this fact to point identify C' on a sequence. Specifically, let us define the set $\mathcal{K} \subset \mathbb{Z}$ and $(y_k)_{k\in\mathcal{K}}$ by induction as follows. First, $0 \in \mathcal{K}$, while y_0 was introduced above. Second, if $k \geq 0$ is such that $y_k \in \mathcal{Y}_1$, then $k + 1 \in \mathcal{K}$ and $y_{k+1} = H(y_k)$. Similarly, if $k \leq 0$ and $y_k \in \mathcal{Y}_2 \equiv \mathrm{Supp}(Y|X = 2)$, then $k - 1 \in \mathcal{K}$ and $y_{k-1} = H^{-1}(y_k)$. Figure 2 illustrates the construction of $(y_k)_{k\in\mathcal{K}}$, in a case where

---

[14]This conclusion is consistent with our own experience. We both worked at Insee in the household survey methodology unit between 2000 and 2003. We are not aware of any particular change related to the interviewers at that time.
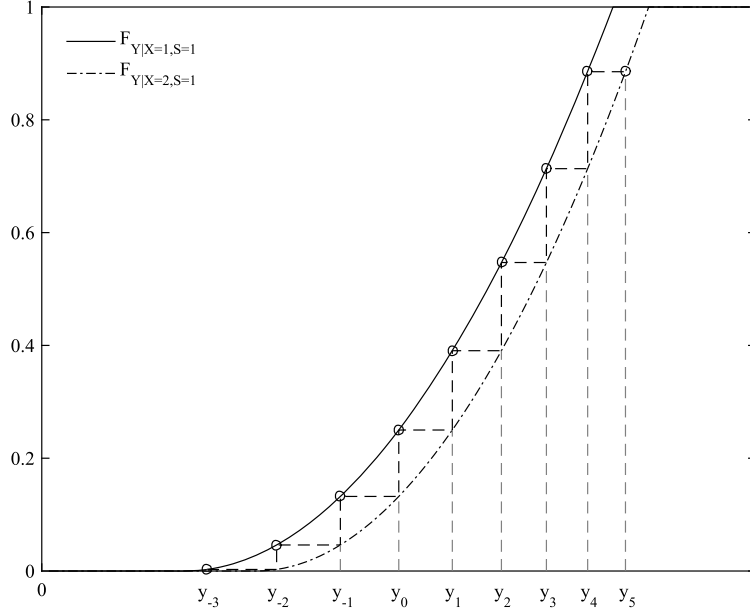
FIGURE 2. Construction of $(y_k)_{k \in \mathcal{K}}$.

$0 < \underline{\theta} < \overline{\theta} < +\infty$. $\mathcal{K}$ is then finite because $\inf \mathcal{Y}_1 < \inf \mathcal{Y}_2$ and $\sup \mathcal{Y}_1 < \sup \mathcal{Y}_2$. Otherwise, $\mathcal{K}$ is infinite, and equal to $\mathbb{Z}$ when $\underline{\theta} = 0$ and $\overline{\theta} = +\infty$.

By an immediate induction based on (3.3), we have, for all $k \in \mathcal{K}$,

$$C'(y_k) = \big(\delta(2)/\delta(1)\big)^k C'(y_0) = \big(\delta(2)/\delta(1)\big)^k c_0.$$

Hence, given c_0, we point identify $C'(y_k)$ for all $k \in \mathcal{K}$. Put it differently, we point identify the ratio $C'(\cdot)/C'(y_0)$ on a sequence of points $(y_k)_{k \in \mathcal{K}}$ determined by the (identified) conditional distributions of Y.

Besides, let $\theta_k = (\delta(1)/\delta(2))^k \delta(1)/c_0$ for all $k \in \mathbb{Z}$. Because $\theta_k = \delta(1)/C'(y_k)$, $\theta_k = \theta(1, y_k)$ for all $k \in \mathcal{K}$. Hence, by strict monotonicity of $\theta(1, \cdot)$,

$$F_{\theta|X=1,S=s}(\theta_k) = 1 - F_{Y|X=1,S=s}(y_k).$$

Because $\theta(2, y_k) = [\delta(2)/\delta(1)]\theta(1, y_k) = \theta_{k-1}$, we also have, for all $k \in \mathbb{Z}$ such that $k + 1 \in \mathcal{K}$,

$$F_{\theta|X=2,S=s}(\theta_k) = 1 - F_{Y|X=2,S=s}(y_{k+1}).$$

We have therefore defined sequences on which C' and $F_{\theta|X,S}$ are point identified. Elsewhere, these functions can be bounded, using monotonicity arguments. Formally, let

$$\underline{C}'(y) = \sup\left\{0, \left(\frac{\delta(2)}{\delta(1)}\right)^k c_0 : k \in \mathcal{K}, y_k \le y\right\},$$

$$\overline{C}'(y) = \inf\left\{+\infty, \left(\frac{\delta(2)}{\delta(1)}\right)^k c_0 : k \in \mathcal{K}, y_k \le y\right\},$$

$$\underline{F}_{\theta|X=x,S=s}(\theta) = \sup\{0, F_{\theta|X=x,S=s}(\theta_k) : k + \mathbb{1}\{x = 2\} \in \mathcal{K}, \theta_k \leq \theta\},$$

$$\overline{F}_{\theta|X=x,S=s}(\theta) = \inf\{1, F_{\theta|X=x,S=s}(\theta_k) : k + \mathbb{1}\{x = 2\} \in \mathcal{K}, \theta_k \geq \theta\}.$$

Theorem 3.2 shows that these are indeed bounds on C' and F_θ, and that they are sharp in a pointwise sense.

THEOREM 3.2. *Suppose that Assumptions 1–5 hold. Then for all* $(x, s, y, \theta) \in \{1, 2\} \times \{0, 1\} \times (0, 1) \times (0, +\infty)$,

$$C'(y) \in \left[\underline{C}'(y), \overline{C}'(y)\right], F_{\theta|X=x,S=s}(\theta) \in \left[\underline{F}_{\theta|X=x,S=s}(\theta), \overline{F}_{\theta|X=x,S=s}(\theta)\right].$$

These bounds are sharp. Moreover, $\underline{C}'(y_k) = \overline{C}'(y_k)$ *for all* $k \in \mathcal{K}$ *and* $\underline{F}_{\theta|X=x,S=s}(\theta_k) = \overline{F}_{\theta|X=x,S=s}(\theta_k)$ *for all* k *such that* $k + \mathbb{1}\{x = 2\} \in \mathcal{K}$. *Thus,* C' *and* $F_{\theta|X=x,S=s}$ *are point identified on these sequences.*

Theorem 3.2 provides the best nonparametric bounds on the agents' cost function and the distribution of heterogeneity of interviewers. Because it is based on agent's program only, the bounds are valid whether or not the contracts are optimal and despite potential selection effects. On the other hand, our identification result strongly relies on the use of an exogenous change. Without variations in the contracts (i.e., when we observe data from only one menu of contract or when the change is endogenous), we prove in the Appendix that the model is not identified. Any increasing marginal cost function C' or any distribution function F_θ can be rationalized by the data.

This result is related to other identification results in the literature. The model can be written equivalently as

$$\ln C'(Y) = \ln \delta(X) + \varepsilon,$$

with $\varepsilon = -\ln\theta$ and $X \perp\!\!\!\perp \varepsilon | S = 1$. Hence, it may be seen as a particular case of a nonparametric transformation models or, equivalently in duration models, of a generalized accelerated failure time (GAFT) model. The identification of such models in the case of a discrete X, has been studied by Abbring and Ridder (2015). They show that if $\underline{\theta} = 0$, $\overline{\theta} = +\infty$ and regularity conditions hold at these boundaries, C' is point identified. They also exhibit counterexamples if such regularity conditions are not met. A first difference between their framework and ours is that here, the support of θ may not be large. We show that in such a case, the model is still partially identified. Second, we show that with a large support but without the regularity conditions imposed by Abbring and Ridder (2015), the model is still partially identified.

Abbring and Ridder's result implies that the model is point identified with a large support on θ, under arguably mild regularity conditions. However, point identification relies very much on the behavior of H at the boundaries, and is therefore expected to be fragile. To illustrate this, suppose that C' is regularly varying at 0, namely $\lim_{y \to 0} C'(ty)/C'(y) = t^\alpha$ for all $t > 0$ and some (unknown) $\alpha \geq 0$. Let also $H^{(1)}(y) = H(y)$

and for any $k \geq 1$, $H^{(k+1)}(y) = H^{(k)}[H(y)]$. Then C' is identified in two steps. First, one can show that α is identified by

$$\alpha = \frac{\ln\big(\delta(2)/\delta(1)\big)}{\ln\Big[\lim_{y \to 0} H(y)/y\Big]}.$$

Second, $C'(y)$ can be shown to be point identified by

$$C'(y) = c_0 \left[\frac{\lim_{k \to \infty} \left(\frac{\delta(1)}{\delta(2)}\right)^k H^{(k)}(y)}{\lim_{k \to \infty} \left(\frac{\delta(1)}{\delta(2)}\right)^k H^{(k)}(y_0)} \right]^{\alpha}.$$

This formula shows that any small variation in H around 0 is amplified when taking the limit, rendering estimation of C' based on this equation very difficult. Related to this issue, no paper has addressed so far the estimation of nonparametric transformation or GAFT models with only discrete regressors.

Theorem 3.2 is also related to Guerre, Perrigne, and Vuong (2009), who show that exogenous changes are necessary but also sufficient to point identify first-price auction models with risk averse bidders. The reason why they obtain point identification rather than partial identification as here is that in their framework, the bidders' strategies cross at the lowest valuation, and this crossing point can be used for identification. In our framework but with other types of contracts, the functions $F_{Y|X=1,S=1}$ and $F_{Y|X=2,S=1}$ could cross inside the support of Y, leading also to point identification (see D'Haultfœuille and Février (2010)).[15]

Finally, our result imply that standard parametric models on C' and $F_{\theta|X=x,S=s}$ are identified with an exogenous change. For instance, the parameters of a lognormal or Weibull distribution are identified thanks to the knowledge of $F_{\theta|X=x,S=s}$ on the sequence $(\theta_k)_{k+\mathbb{1}\{x=2\}\in\mathcal{K}}$. Actually, because we retrieve an infinite sequence of points on C' and $F_{\theta|X,S}$, such standard parametric models are overidentified. The sequences $(C'(y_k))_{k\in\mathcal{K}}$ and $(F_{\theta|X=x,S=s}(\theta_k))_{k+\mathbb{1}\{x=2\}\in\mathcal{K}}$ may thus serve as a guidance for choosing appropriate parametric restrictions, as will be the case in Section 4.3 below.

3.3 *Nonparametric estimation of the cost function and the distribution of types*

We now turn to the nonparametric estimation of C' and $F_{\theta|X,S}$. In particular, we study the behavior of these estimators when the number of interviewers tend to infinity in the sense that $L \equiv \min_{(x,s)\in\{1,2\}\times\{0,1\}} \#\{i : X_i = x, S_i = s\} \to \infty$. We impose hereafter the following standard assumption of independent sampling.

---

[15] Another solution to recover point identification would be to use the principal's program together with restrictions on its objective function. In our framework yet, this program has no identification power on C' and $F_{\theta|X,S|}$, but simply allows us to recover its objective function. See Section 4.2 below for more details.

ASSUMPTION 6 (Independent sampling). *For any $(x, s) \in \{1, 2\} \times \{0, 1\}$, the sample $(\theta_i, R_i, N_i)_{i:X_i=x,S_i=s}$ is made of i.i.d. vectors.*

Our nonparametric estimation method follows closely the identification strategy and may be decomposed into two steps. We first estimate the conditional distribution $F_{Y|X,S}$ of the unobserved probabilities. We then estimate bounds on the primitive functions C' and $F_{\theta|X,S}$, using the result of Theorem 3.2.

For the first step, we use a sieve maximum likelihood estimator (see, e.g., Chen (2007), for a survey on sieve estimation). We choose to approximate the densities[16] $f_{Y|X,S}$ by functions of the sieve space

$$\mathcal{F}_L = \left\{ f : 0 \leq f \leq M \ln K_L, \int_0^1 f(x)\,dx = 1 \text{ and } \sqrt{f} \in \mathcal{P}_{K_L} \right\},$$

where $\mathcal{P}_J$ denotes the space of polynomials of order at most J, M is a constant and $(K_L)_{L \in \mathbb{N}}$ is an increasing sequence tending to infinity. We thus approximate the conditional density $f_{Y|X,S}$ by squares of polynomials that integrate to one. Squares of polynomials are convenient because they ensure that the estimated density is positive, are easy to integrate, and lead to a simple likelihood.[17] To see this, let us consider $f(\cdot; \mathbf{a}) \in \mathcal{F}_L$ defined by

$$f(x; \mathbf{a}) = \left(\sum_{k=0}^{K_L} a_k x^k \right)^2 \equiv \sum_{k=0}^{2K_L} b_k(\mathbf{a}) x^k,$$

where $\mathbf{a} = (a_0, \ldots, a_{K_L})$ and $b_k(\mathbf{a}) = \sum_{\ell=\max(0, k-K_L)}^{\min(k, K_L)} a_\ell a_{k-\ell}$. The likelihood of an observation corresponding to $f(\cdot; \mathbf{a})$ is, by independence between Y and N conditional on $X = x, S = s$,

$$\Pr(R = r | N = n, X = x, S = s) = E\left[\Pr(R = r | N = n, Y, X = x, S = s) | X = x, S = s \right]$$

$$= \binom{r}{n} E\left[y(x, \theta)^r (1 - y(x, \theta))^{n-r} | X = x, S = s \right]$$

$$= \binom{r}{n} \int_0^1 \sum_{k=0}^{2K_L} b_k(\mathbf{a}) y^{r+k} (1 - y)^{n-r}\,dy$$

$$= \binom{r}{n} \sum_{k=0}^{2K_L} b_k(\mathbf{a}) B(r + k + 1, n - r + 1), \qquad (3.4)$$

---

[16] Assumption 2 and the equality $F_{\theta|X=x,S=s}(\delta(x)/C'(y)) = 1 - F_{Y|X=x,S=s}(y)$ ensure that the density of Y conditional on X, S does exist.

[17] We also restrict ourselves to bounded polynomials. This ensures that $\mathcal{F}_L$ is compact and simplifies the consistency proof.

where $B(\cdot, \cdot)$ denotes the beta function. We let $\widehat{f}_{Y|X,S}$ denote the maximum likelihood estimator (over $\mathcal{F}_L$) of $f_{Y|X=x,S=s}$.[18] We then estimate $F_{Y|X,S}$ and $F_{Y|X,S}^{-1}$ by $\widehat{F}_{Y|X,S}(y) = \int_0^y \widehat{f}_{Y|X,S}(u)\,du$ and $\widehat{F_{Y|X,S}^{-1}}(u) = \widehat{F}_{Y|X,S}^{-1}(x)$.

We now turn to the estimation of C' and $F_{\theta|X,S}$. First, we estimate H and $\mathcal{Y}_x(x \in \{1, 2\})$ by respectively $\widehat{H}(y) = \widehat{F}_{Y|X=2,S=1}^{-1} \circ \widehat{F}_{Y|X=1,S=1}(y)$ and $\widehat{\mathcal{Y}}_x = [\widehat{F}_{Y|X=x,S=1}^{-1}(\tau_L),$ $\widehat{F}_{Y|X=x,S=1}^{-1}(1 - \tau_L)]$, for a sequence τ_L tending to 0. Second, we define $\widehat{\mathcal{K}}$ and $(\widehat{y}_k)_{k \in \widehat{\mathcal{K}}}$ as before. $0 \in \widehat{\mathcal{K}}$ and if $k \geq 0$ is such that $\widehat{y}_k \in \widehat{\mathcal{Y}}_1$, then $k + 1 \in \widehat{\mathcal{K}}$ and $\widehat{y}_{k+1} = \widehat{H}(\widehat{y}_k)$. Similarly, if $k \leq 0$ and $\widehat{y}_k \in \widehat{\mathcal{Y}}_2$, then $k - 1 \in \widehat{\mathcal{K}}$ and $\widehat{y}_{k-1} = \widehat{H}^{-1}(\widehat{y}_k)$. Note that θ_k does not need to be estimated. Then we consider plug-in estimators for the bounds on C' and $F_{\theta|X,S}$:

$$\underline{\widehat{C}}'(y) = \sup\{0, (\delta(2)/\delta(1))^k c_0 : k \in \widehat{\mathcal{K}}, \widehat{y}_k \leq y\},$$

$$\overline{\widehat{C}}'(y) = \inf\{+\infty, (\delta(2)/\delta(1))^k c_0 : k \in \widehat{\mathcal{K}}, \widehat{y}_k \geq y\},$$

$$\underline{\widehat{F}}_{\theta|X=x,S=s}(\theta) = \sup\{0, 1 - \widehat{F}_{Y|X=x,S=s}(\widehat{y}_k) : k + \mathbb{1}\{x = 2\} \in \widehat{\mathcal{K}}, \theta_k \leq \theta\},$$

$$\overline{\widehat{F}}_{\theta|X=x,S=s}(\theta) = \inf\{1, 1 - \widehat{F}_{Y|X=x,S=s}(\widehat{y}_k) : k + \mathbb{1}\{x = 2\} \in \widehat{\mathcal{K}}, \theta_k \geq \theta\}.$$

Theorem 3.3 below establishes the consistency of these bounds under the following regularity conditions.

ASSUMPTION 7. *For all* $k \in \mathcal{K}$, $y_k \notin \{\inf \mathcal{Y}_2, \sup \mathcal{Y}_1\}$. *For all* $(x, s) \in \{1, 2\} \times \{0, 1\}$, $\lim_{\theta \to \overline{\theta}} \theta^2 f_{\theta|X=x,S=s}(\theta) = 0$ *and* C'' *is bounded on* $(0, 1/2)$. *Either* $\underline{\theta} = 0$ *and* $\lim_{y \to 1} \frac{C''(y)}{C'(y)^2}$ *exists and is finite, or* $\underline{\theta} > 0$ *and* $f_{\theta|X=x,S=s}(\underline{\theta}) = 0$. *For all* $u > 0$, $E(u^N|X = x, S = s) < \infty$.

The condition on the sequence hold automatically when $\text{Supp}(\theta) = \mathbb{R}^+$. Otherwise, we just impose that the smallest (resp., largest) value of y_k is not equal to $\inf \mathcal{Y}_2$ (resp., $\sup \mathcal{Y}_1$), which is a very mild restriction on the choice of y_0. The conditions on $f_{\theta|X=x,S=s}$ and C' ensure that $f_{Y|X=x,S=s}$ is continuous on $[0, 1]$, whether or not $1/\underline{\theta}$ and $\overline{\theta}$ are finite. The last condition imposes light tails for the conditional distribution of N.

THEOREM 3.3. *Suppose that Assumptions 1–7 hold,* $K_L \to \infty$ *and* $K_L^2 \ln K_L/L \to 0$. *Then* $\widehat{F}_{Y|X=x,S=s}$ *is uniformly consistent on* $[0, 1]$. *Moreover, for any sequence* $\tau_L \to 0$ *such that* $P(\sup_{y \in [0,1]} |\widehat{F}_{Y|X=x,S=s}(y) - F_{Y|X=x,S=s}(y)| < \tau_L) \to 1$, $\underline{\widehat{F}}_{\theta|X,S}(\theta)$ *and* $\overline{\widehat{F}}_{\theta|X,S}(\theta)$ *are consistent for all* $\theta > 0$. $\underline{\widehat{C}}'(y)$ *and* $\overline{\widehat{C}}'(y)$ *are consistent on every* $y \in (0, 1)\backslash\{y_k, k \in \mathcal{K}\backslash\{0\}\}$. *Finally, for all* $k \in \mathcal{K}$,

$$\left(\widehat{y}_k, \overline{\widehat{C}}'(\widehat{y}_k) = \underline{\widehat{C}}'(\widehat{y}_k)\right) \xrightarrow{\mathbb{P}} \left(y_k, C'(y_k)\right).$$

Theorem 3.3 has four parts. The first establishes the uniform consistency of the nonparametric estimator of $F_{Y|X=x,S=s}$. The second shows that the estimated bounds on

---

[18] Intuitively, this estimator weights more interviewers with a large N, because for them, the distribution R/N is closer to the one of Y: by the central limit theorem, we have approximately $R/N = Y + \sqrt{Y(1 - Y)/N}\varepsilon$, with $\varepsilon|N, Y \sim \mathcal{N}(0, 1)$. Because of the error term $\sqrt{Y(1 - Y)/N}\varepsilon$, it is more difficult to discriminate between two parametric distributions on Y when N is small.

$F_{\theta|X,S}$ are consistent. The third shows the convergence of $\underline{C}'$ and $\overline{C}'$ outside the sequence $(y_k)_{k\in\mathcal{K}}$. Even if consistency fails in general on this sequence, the last part of the theorem shows point consistency in $\mathbb{R}^2$ of the estimated sequence $(\widehat{y}_k, \widehat{\overline{C}'}(\widehat{y}_k))$. As a consequence C' and $F_{\theta|X,S}$ are well estimated on the sequences where they are point identified, while sharp bounds are consistently recovered anywhere else.

Consistency of the bounds requires to choose τ_L appropriately, which is difficult because the rate of convergence of $\widehat{F}_{Y|X=x,S=s}$ is unknown. But interestingly, if we consider τ_L fixed, independent of L, one can show that we still estimate consistently the bounds on C' and $F_{\theta|X,S}$, but only on subsets of $(0, 1)$ and $\mathbb{R}^+$, respectively. Elsewhere, we get outer bounds on these functions because, basically, $\widehat{\mathcal{K}} \subsetneq \mathcal{K}$ with probability approaching one. Therefore, letting τ_L tend to zero at the appropriate rate is required only to estimate optimal bounds everywhere.

3.4 *Results*

We estimate in a first step $F_{Y|X=x,S=s}$ by the sieve MLE proposed above. As usual, there is a trade-off between bias and variance in the choice of K_L. Empirically, the estimates do not seem to be too smooth or too erratic for K_L between 3 and 6. Results are quite similar in this range, and we choose $K_L = 3$ for the stayers and $K_L = 2$ for the movers. The corresponding estimates are displayed in Figure 3. As predicted by the theory, the distribution function of $y(2, \theta)$ for stayers dominates stochastically the one of $y(1, \theta)$ on most part of $(0, 1)$ (see the left graph). We also observe that for movers, the estimated distribution of $y(2, \theta)$ dominates stochastically the one of $y(1, \theta)$ (see the right graph). This arises because of incentive effects but also possibly because of selection effects. We discuss in the next section the existence of selection effects in our context.

We now estimate nonparametrically the sharp bounds on $F_{\theta|X=2}$ and C'. $F_{\theta|X=2}$ is interesting as it corresponds to the distribution of interviewers' types on the 2003 survey. We obtain similar patterns for $F_{\theta|X=1}$. We first choose a starting value y_0 close to the median of $\widehat{F}_{Y|X=1,S=1}$, namely $y_0 = 0.8$, in order to get more precise estimates for central
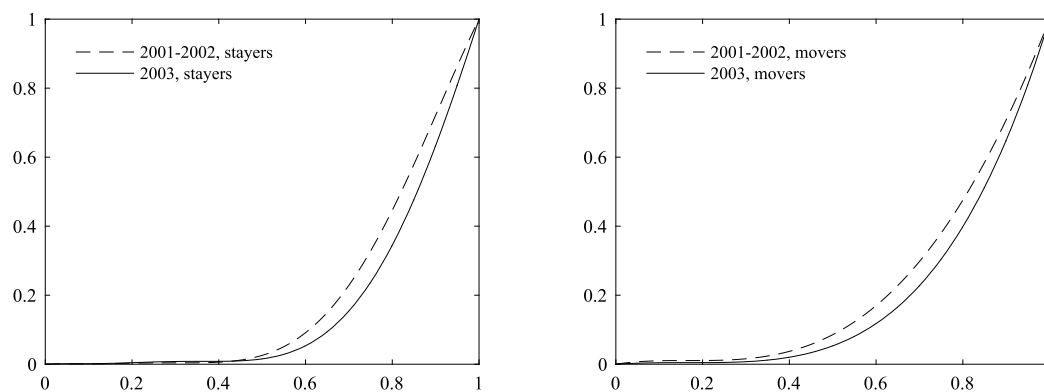

FIGURE 3. Sieve MLE estimates of $F_{Y|X,S}$. *Notes*: 161 observations for $X = 1, S = 0$, 79 for $X = 2, S = 0$, and 374 for $S = 1$.
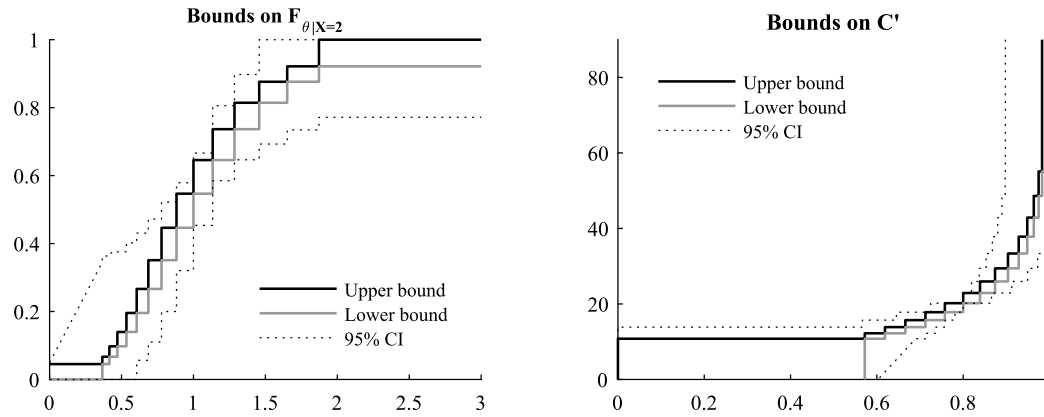
FIGURE 4. Estimated bounds on $F_{\theta|X=2}$ and C'. *Notes*: the 95% confidence intervals are computed by bootstrap.

values of $F_{\theta|X=2}$ and C'.[19] For that y_0, we impose the normalization $C'(y_0) = \delta(1)$, which is equivalent to imposing $\theta(1, y_0) = 1$. We then choose $\tau_L = 0.05$, which leads to estimating 14 points in $\widehat{\mathcal{K}}$. Of course, other choices of τ_L enlarge or shrink $\widehat{\mathcal{K}}$. For instance, with $\tau_L = 0.01$ and $\tau_L = 0.10$, we get respectively 22 and 11 points. But the bounds on the functions are not altered substantially; they only change for large values where standard errors are large anyway.

Figure 4 displays the estimates of the bounds on $F_{\theta|X=2}$ and C', and their 95% confidence interval obtained by bootstrap. 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 significantly lower for higher ones. Finally, the width of the confidence intervals on the bounds of $F_{\theta|X=2}$ (resp., C') increases with $|\theta - 1|$ (resp., $|y - 0.8|$), reflecting the fact that, as expected, the estimation error increases with $|k|$.

3.5 *Tests of the model and robustness checks*

The results above rely on a few assumptions that we now check. The first part of Assumption 3 implies that $y(n, x, \theta)$ does not depend on n, so that Y is only a function of (X, θ). Then, by the second part of Assumption 3,

$$N \perp\!\!\!\perp Y | X, S. \tag{3.5}$$

This condition is testable. To see this, note that by definition of Y, $E(R|N, Y, X, S|) = NY$. Then

$$E(R/N|N, X, S) = E(Y|N, X, S) = E[Y|X, S]. \tag{3.6}$$

---

[19]We have checked that other values of y_0 do not modify the choice of the parametric families that is made using our nonparametric estimates.

In other words, R/N is mean independent of N conditional on X, S. We test the restriction (3.6) by considering the model

$$R/N = \zeta(X, S) + g(N) + \varepsilon,$$

where $E(\varepsilon|N, S) = 0$, ζ is a year times participation status fixed effect and we distinguish in X between the 2001 and 2002 survey to be robust to the exogeneity condition (Assumption 5). Equation (3.6) implies that the function g should be equal to zero. We perform a test of this restriction using linear, quadratic, and flexible parametric specifications for g. We consider for this latter specification the piecewise linear function $g(n) = g_1 n + g_2 (n - 10)^+ + g_3 (n - 20)^+$ (with $x^+ = \max(0, x)$), which can detect more complex dependence between R/N and N under the alternative.[20]

Results are presented in Table 3. We accept the null hypothesis that $g = 0$ at standard levels in any of the three specification. Moreover, the point estimates are very small. The results of the linear specification for instance imply that a very substantial increase of 10 households to interview is associated to a small increase of one percentage point in the average response rate.

Next, our results crucially hinge upon Assumption 5. We present hereafter three suggestive tests of $X \perp\!\!\!\perp \theta|S = 1$. The idea behind the first is that households are more or less difficult to contact depending on their characteristics. If the distribution of their characteristics changed in 2003, this could induce a shift in the distribution of θ, thus violating Assumption 5. We thus check whether the average characteristics of the housings attributed to stayers differ systematically between 2001–2002 and 2003. We can only use housings' characteristics that are available for both respondents and nonrespondents.

TABLE 3. Test of Assumption 3 based on regressions of response rates on functions g of subsample sizes.

Variable	Linear g	Quadratic g	Piecewise Linear g
Subsample size	0.001 (0.001)	0.0028 (0.0028)	0.0045 (0.0032)
Subsample size squared	–	−0.0001 (0.0001)	–
(Subsample size − 10)$^+$	–	–	−0.0058 (0.0043)
(Subsample size − 20)$^+$	–	–	0.0026 (0.0031)
Participation status × year included	Yes	Yes	Yes
R^2	0.014	0.014	0.016
p-value of the test $g = 0$	0.29	0.54	0.52

Note: 984 observations. We control for participation status interacted with the year. Standard errors are clustered by interviewers to take into account the dependence arising because of θ_i.

---

[20]We consider in Appendix B.1 another test using not only the first moment of R, but its whole distribution. However, this test also relies on Assumption 1, and thus does not solely test Assumption 3. Again, we fail to reject the null at standard level with this alternative test.

TABLE 4. Stability test of housing characteristics (p-values).

Variable	All Stayers	Stayers With exp > 10
Proportion of collective housings	0.21	0.20
Average number of rooms	0.99	0.93
Proportion of housings in rural areas	0.11	0.12
Proportion of housings in small towns	0.28	0.49
Proportion of housings in large towns	0.83	0.93
Number of observations	748	280

Note: p-values of the Kolmogorov–Smirnov test that the distribution of proportions of, for example, individual housing, remains constant between 2001–2002 and 2003.

We use hereafter variables that are known to be correlated with nonresponse, namely the dummy of being in a collective housing, the number of rooms and the dummies of being in rural areas, in small towns (of less than 100,000 inhabitants), and in large towns (of more than 100,000 inhabitants). We then perform a Kolmogorov–Smirnov test that the distributions of the averages over stayers of these variables did not change between 2001–2002 and 2003. The results are displayed in the first column of Table 4. No test is significant at the 10% level, which clearly supports our assumption.

Our second test aims at testing whether areas could have been attributed to interviewers differently in 2003, as a function of the areas' and the interviewers' characteristics. A change in the "match" between interviewers and areas could change the distribution of θ, even though both marginal distributions (of interviewers and areas) have remained the same. To test for this issue, we investigate whether experienced interviewers had a different distribution of housing characteristics in 2003 than in 2001–2002. We rely on the same Kolmogorov–Smirnov tests as above, but now focusing only on interviewers' with more than 10 years of experience (results are similar with other thresholds). The results, displayed in the second column of Table 4, support again our claim that the assignment of areas to interviewers was not different in 2003.

A final concern on the assumption that $X \perp\!\!\!\perp \theta|S = 1$ is related to experience: stayers have accumulated experience between 2001 or 2002 and 2003, and could therefore be more efficient in 2003. This learning-by-doing effect is unlikely to be of first order here since interviewers have already 9 years of experience on average. Yet, we can test for this possibility under some assumptions. Let E denote the experience of an interviewer and suppose that $\partial C/\partial y(y; E) = \beta(E))(y/(1 - y))^\zeta$. Combined with Assumption 1, this implies that the dummy y_{ixk} of whether interviewer i managed or not to interview household k in survey $x \in \{0, 1, 2\}$ satisfies

$$y_{ixk} = \mathbb{1}\big\{\gamma\mathbb{1}\{x = 2\} + \widetilde{\beta}(E_{ix}) + \widetilde{\theta}_i + \varepsilon_{ixk} \geq 0\big\}, \tag{3.7}$$

where $\gamma = \ln \delta(2)/\zeta$, $\widetilde{\beta}(E_{ix}) = -\ln(\beta(E_{ix}))$, $\widetilde{\theta}_i = -\frac{1}{\zeta}\ln(\theta_i)$ and the $(\varepsilon_{ijk})_{i,j,k}$ are independent and follow a logit distribution. We recall that γ measures the incentive effect of 2003 with respect to 2001–2002, and is therefore key in our analysis. Hence, we want

TABLE 5. Effects of interviewers' experience versus incentive effects.

Specification of $\beta(E)$	Estimate of γ	Estimate of b	p-Value of No Effect of Experience
$\beta(E) = 1$	0.24 (0.06)	–	–
$\beta(E) = E^{-b}$	0.26 (0.07)	−0.09 (0.12)	0.44
$\beta(E) = 1 + (\exp(-b) - 1)\mathbb{1}\{E \geq 3\}$	0.25 (0.06)	−0.08 (0.15)	0.60
$\beta(E) = 1 + (\exp(-b) - 1)\mathbb{1}\{E \geq 5\}$	0.24 (0.06)	0.05 (0.23)	0.84

Note: Estimates of b_1 in the fixed-effect logit model (3.7), with various parametrizations of $\beta(E)$. Standard errors under parentheses. 9851 observations.

to check whether the estimate of γ is sensitive to the introduction of the effect of experience. We estimate for that purpose Model (3.7) using various parametric specifications for $\beta(E_{ix})$. Given that $E_{ix} = E_i + x$, we cannot identify flexible functions β, since $\widetilde{\beta}(E_{ix})$ would become collinear with $\mathbb{1}\{x = 2\}$. For instance, we cannot identify (b_1, b_2) if we let $\beta(E) = \exp(b_1 E + b_2 E^2)$. We consider hereafter three specifications: $\beta(E) = E^{-b}$, $\beta(E) = 1 + (\exp(-b) - 1)\mathbb{1}\{E \geq 3\}$ and $\beta(E) = 1 + (\exp(-b) - 1)\mathbb{1}\{E \geq 5\}$.

The results are displayed in Table 5. We obtain two conclusions. First, the coefficient of γ is hardly affected by the introduction of experience. Second, experience has no significant effect in the three specifications we consider on β. Hence, this test suggests that the effect of experience would threaten our conclusion.

Finally, when combining Assumptions 1, 3, and the polynomial restriction on $f_{Y|X,S}$ behind the sieve MLE, we obtain a relatively parsimonious parametric model for the distribution of R conditional on N. Specifically, the probabilities $\Pr(R = r | N = n, X = x, S = s)$ only depend on $K_L + 1$ parameters (see equation (3.4)). Hence, if seen as a parametric model (where K_L is fixed as L tends to infinity), the model is largely overidentified, as there are many more moment conditions corresponding to all the equalities implied by equation (3.4), than parameters. To assess whether this model fits well the data, and thus whether the polynomial restriction of $f_{Y|X,S}$ is reasonable (under the maintained Assumptions 1 and 3), we consider a GMM overidentification test, for each of the four subpopulations $\{X = x, S = s\}$, $(x, s) \in \{0, 1\}^2$.

Given the moderate subsample sizes and the large number of moment conditions, we can expect the quantiles of the asymptotic distribution of this test to underestimate the true quantiles under the null. To conduct more reliable inference, we thus use the bootstrap instead.[21] We compute bootstrap critical values by first drawing with replacement the $(N_i)_{i:X_i=x,S_i=s}$ and then drawing $R|N_i = n$ using (3.4), with **a** replaced by the sieve MLE estimator. The p-values of the tests are displayed in Table 6. For the four subpopulations $\{X = x, S = s\}$, $(x, s) \in \{0, 1\}^2$, we fail to reject the null hypothesis at all usual levels. This suggests that under the maintained Assumptions 1 and 3, the polynomial restriction we rely on in the sieve MLE is reasonable.

---

[21]When using the asymptotic distribution instead of the bootstrap, we indeed obtain much higher p-values.

TABLE 6. Overidentification tests of Assumption 3 and the polynomial restriction on $f_{Y|X,S}$.

	Stayers 2001–2002	Stayers 2003	Movers 2001–2002	Movers 2003
p-value	0.70	0.23	0.74	0.48
Number of obs.	348	374	161	79

Note: GMM overidentification test of (3.4), with $K_L = 3$ for the stayers and $K_L = 2$ for the movers. The optimal weighting matrix is estimated using the sieve MLE estimator of **a** and the critical values are estimated by bootstrap.

4. POLICY ANALYSIS

In this section, we compare the current contracts with the optimal, nonlinear ones, and with settings without asymmetries of information. The information that we have recovered so far on the interviewers' type and their cost function is used for that purpose. But before performing this analysis, we have to check that there is no selection effects in our context. If these potential effects were not an issue for identifying the interviewers' utility function, they would complicate substantially the policy analysis because basically, different contracts would select different types of interviewers.

4.1 *Testing the absence of selection effects*

To evaluate the average response rate that would prevail under contracts that differ from the actual ones, we have to take into account selection effects, namely that more attractive contracts may attract betters interviewers, for instance. We provide here statistical evidence that this is likely not the case in our context. More precisely, we test whether the distribution of movers of the 2001–2002 surveys are identical to the one of the movers of the 2003 survey (while the distribution of stayers could differ from them). Formally, this amounts to test $H_0 : \theta \perp\!\!\!\perp X|S = 0$. To perform such a test, remark that under H_0, we have, for $s \in \{0, 1\}$,

$$F_{Y|X=1,S=s}(y) = F_{\theta|X=1,S=s}\big[\theta(1, y)\big]$$
$$= F_{\theta|X=2,S=s}\big[\theta\big(2, y\big(2, \theta(1, y)\big)\big)\big]$$
$$= F_{Y|X=2,S=s}\big(y\big(2, \theta(1, y)\big)\big).$$

This shows that under H_0, $F_{Y|X=2,S=s}^{-1} \circ F_{Y|X=1,S=s}$ does not depend on s. As a result,

$$F_{Y|X=2,S=0} = F_{Y|X=1,S=0} \circ H^{-1}. \tag{4.1}$$

In other words, if we transform the distribution of the probabilities chosen by the 2002 movers using the quantile-quantile transform of the stayers, we should obtain the distribution of the probabilities chosen by the 2003 movers. Equation (4.1) holds on the domain of definition $\mathcal{Y}_2$ of H^{-1}, but by letting $H^{-1}(y) = 0$ for $y < \inf \mathcal{Y}_2$ and $H^{-1}(y) = 1$ for $y > \sup \mathcal{Y}_2$, (4.1) actually holds on the whole interval $[0, 1]$.

Equation (4.1) suggests the use of the test statistic $T = \sup_{y \in [0,1]} |\widehat{\Delta}(y)|$, where $\widehat{\Delta}$ is the nonparametric estimator of $\Delta = F_{Y|X=2,S=0} - F_{Y|X=1,S=0} \circ H^{-1}$. The logic behind this

test statistic is that under the null hypothesis H_0, $\Delta = 0$. The main challenge here is to derive the distribution of T under H_0. We estimate this distribution by the distribution of T^*, the test statistic of bootstrap samples drawn under H_0. To draw under H_0, we consider estimators of the distributions of R conditional on (N, X, S) that satisfy H_0 and are consistent under this hypothesis:

1. For the stayers and the 2001–2002 movers, we first draw N from its empirical distribution, and independently of N, Y according to the sieve MLE estimator $\widehat{F}_{Y|X=x,S=s}$. We then draw $R|N, Y \sim \text{Binomial}(N, Y)$.

2. For the 2003 movers, we draw N from its empirical distribution, and independently of N, Y according to $\widehat{F}_{Y|X=1,S=0} \circ \widehat{H}^{-1}$. We then draw $R|N, Y \sim \text{Binomial}(N, Y)$.

Our estimators of $F_{Y|X=x,S=s}$ are consistent by Theorem 3.3. Moreover, the bootstrap distribution corresponding to $F_{Y|X=2,S=0}$ satisfies the null hypothesis by construction. Thus, the bootstrap distribution we consider is consistent under the null hypothesis.

   We obtain $T \simeq 0.038$ and a p-value of 0.88, and thus cannot reject the absence of selection effects. This result may seem surprising, given the importance of selection effects obtained by, for example, Lazear (2000). This difference may stem from the pattern in workers' turnover. Whereas new workers were hired by the car glass company in Lazear's application, Insee always relies on the same pool of interviewers. Thus, selection effects could only occur through a reallocation of interviewers among this pool. The result of our test suggests that such reallocations are not related to interviewers' productivity.

4.2 *Insee's program and counterfactual contracts*

4.2.1 *Insee's program*   Turning to Insee's program, we suppose that Insee values each interview in survey x as $\lambda(x)$. $\lambda(x)$ represents the "price" of the information contained in a household's answers. The dependence in x reflect the fact that surveys may differ in the "social value" of the information that can be recovered from it. The 2003 survey on education may have been considered by Insee more important than the other ones, as there was much debate at that time 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.

   We suppose that Insee maximizes its objective function by choosing among linear contracts only. The rationale for this assumption is that Insee uses linear contracts for all its household surveys, not only the PCV ones. This feature seems too peculiar to assume that Insee maximizes its objective function among all contracts. Note however that the linear contract chosen by Insee could well be optimal among the larger set of nonlinear contracts defined below. One of our aims is to evaluate the loss by Insee due to restricting to linear contracts, keeping in mind that there could actually be no loss.

   On a related note, Insee also violates the Informativeness Principle, which states that all factors correlated with performance should be included in the contracts (Prendergast

(1999)). For instance, the bonus does not depend on the type of area in which interviewers are working, even if the average response rate in large urban areas (79.8%) is well below the one elsewhere (85.1%). Similarly, the average response rate of Paris area (74.7% in 2003) is significantly lower than the one of the rest of France (84.3%). It may even be the case that Insee observes the type of each interviewers, at least for interviewers hired for several years. Because it does not use this information when proposing its contracts, adverse selection occurs de facto, whether or not Insee observes these types.

Conditional on the participation of the interviewer, the expected profit of Insee in survey x and for a household associated to an interviewer of type θ is therefore $(\lambda(x) - \delta)y(x, \theta, \delta) - w$ when the bonus is δ and the basic wage is fixed to w. Because the basic wage remains almost the same for the three surveys, whereas the bonus increases in the third survey, we suppose that due to constraints, $w(x)$ is fixed so as to ensure that the worse type participates. Regarding $\delta(x)$, we obtain, by optimality of the observed payment scheme among linear contracts and aggregating over all types,

$$\delta(x) = \arg\max_{\delta} E\big[(\lambda(x) - \delta)y(x, \theta, \delta)\big]. \tag{4.2}$$

This program incorporates interviewers' incentive constraint since the probability of response they choose, $y(x, \theta, \delta)$, depends on δ. Also, this program implies that Insee knows the distribution of θ. This standard assumption is credible here. First, Insee knows its interviewers for a long period of time (their average experience in 2001 is 10.5 years). Second, Insee has done many household surveys for a long time, including many PCV surveys before 2001. So it seems reasonable to assume that they know the distribution of the propensity to respond of households.[22]

Program (4.2) implies that $\delta(x)$ satisfies the first-order condition:

$$-E\big[y(x, \theta, \delta(x))\big] + (\lambda(x) - \delta(x))E\left[\frac{\partial y}{\partial \delta}(x, \theta, \delta(x))\right] = 0. \tag{4.3}$$

Under our assumptions above, this first-order condition can be shown to be both necessary and sufficient. Finally, given its policy, Insee's expected current surplus for one household in survey x is

$$\Pi(x) = (\lambda(x) - \delta(x))E\big[y(x, \theta, \delta(x))\big] - w(x). \tag{4.4}$$

4.2.2 *Counterfactual contracts*   The surplus $\Pi(x)$ may not be optimal since Insee restricts itself to linear contracts only, and does not make its contract depend on n. The optimal menu of contract $t_n^*(x)$ for interviewers with n interviews to conduct takes the form of a vector $(t_{n0}^*(x), \ldots, t_{nn}^*(x))$, where $t_{nk}^*(x)$ is the optimal payment of Insee to an interviewer who makes k out of n possible interviews. To define this optimal menu, we

---

[22] Another issue that could invalidate (4.2) is a budget constraint on the average spending per survey. Then $\delta(j)$ could be a corner solution of a constrained version of (4.2), turning (4.3) into an inequality. If the budget constraint for the PCV survey was stable over time, which seems very plausible, the bonus increase in 2003 means however that the bonus did not reach its maximal value in 2001–2002. Still, the 2003 bonus could be a corner solution, which would invalidate the counterfactuals we consider below for 2003. But we obtain very similar results for 2001–2002, so again, it seems unlikely that this issue really affects our results.

first maintain the assumption that Insee ensures universal participation. This implies that $t_{n0}^*(x) = nw(x)$. Second, let $y_n(x, \theta, t)$ be the probability chosen by an interviewer of type θ with n households to interview, when facing the contract $t = (t_0, \ldots, t_n)$. $y_n(x, \theta, t)$ satisfies

$$y_n(x, \theta, t) = \arg \max_{y \in [0,1]} \sum_{k=0}^{n} \binom{n}{k} y^k (1-y)^{n-k} t_k - n\theta C(y). \tag{4.5}$$

The average surplus Insee obtains from an interviewer of type θ, when fixing the transfer vector to t is $\lambda(x) y_n(x, \theta, t)$. The average cost for Insee is

$$\sum_{k=0}^{n} \binom{n}{k} y_n(x, \theta, t)^k (1 - y_n(x, \theta, t))^{n-k} t_k.$$

Therefore, the optimal menu of contract $t_n^*(x)$ satisfies

$$t_n^*(x) = \arg \max_{t \in \{nw(x)\} \times \mathbb{R}^n} E \Bigg[\lambda(x) n y_n(x, \theta, t)$$
$$- \sum_{k=0}^{n} \binom{n}{k} y_n(x, \theta, t)^k (1 - y_n(x, \theta, t))^{n-k} t_k \Bigg]. \tag{4.6}$$

We can also compare the previous surpluses with the one Insee would obtain in two countefactual scenarios. The first is the absence of moral hazard. Insee would then observe the probability chosen by the interviewers, and could implement contracts based on these probabilities. If Insee provided a contract $t_n(y)$ for n households to interview, interviewers of type θ would choose the probability $y_n^{WM}(\theta, t)$ given by the first-order condition

$$t_n'\big(y_n^{WM}(\theta, t_n)\big) = n\theta C'\big(y_n^{WM}(\theta, t_n)\big). \tag{4.7}$$

Then the optimal contract $t_n^{WM}(x, \cdot)$ Insee would choose without moral hazard satisfies

$$t_n^{WM}(x, \cdot) = \arg \max_{t_n(\cdot) : t_n(0) = nw} E\big[\lambda(x) n y_n^{WM}(\theta, t_n) - t_n\big(y_n^{WM}(\theta, t_n)\big)\big]. \tag{4.8}$$

The difference between $t_n^{WM}(x, \cdot)$ and $t_n^*(x)$ is that in the latter case, moral hazard prevents Insee from considering transfers based on the probability of interview. Only transfers based on R (for any fixed n) are possible. But as one can see in (4.6), the situation is actually as if Insee could define contracts based on the probability y chosen by interviewers, but was constrained to use polynomials of order n of y instead of considering any function of y. So we expect the surplus gain from $t_n^*(x)$ to $t_n^{WM}(x, \cdot)$ to be small if optimal polynomial contracts provide a close approximation to unconstrained, optimal contracts.

Finally, we can compute the surplus Insee would obtain without asymmetric information, that is, if it observed and used the type of each interviewer. Under complete information, Insee would be able to fix the probability with which each interviewer would

interview his households. These optimal probabilities $y^C(x, \theta)$ satisfy

$$\lambda(x) = \theta C'\big(y^C(x, \theta)\big). \tag{4.9}$$

Moreover, Insee would recover all the rent from the interviewers. As a result, the optimal transfer function $t^C(x, \cdot)$ with n households to interview is defined by

$$t^C\big(x, y^C(x, \theta)\big) = n\big[\theta C\big(y^C(x, \theta)\big) + w(x)\big], \tag{4.10}$$

where we have used here the normalization $C(0) = 0$.

Finally, to analyze further the role of asymmetric information, it is possible to compare the counterfactual surpluses corresponding to these scenarios with the ones that Insee would obtain if it incorporated some information at its disposal. Still relying on simple linear contracts, Insee could offer, for instance, different contracts in large urban areas versus other areas. Such contracts are given by equation (4.2), where the expectations are taken on the considered populations of interviewers.

4.2.3 *Identification of current and counterfactual surpluses*　The current and counterfactual surpluses are all related to $\lambda(x)$, so not surprisingly, the issue of their identification boils down to whether $\lambda(x)$ is identified or not. Now, by (4.3) and the equality $\partial y / \partial \delta(\theta, \delta) = 1/[\theta C''(y(\theta, \delta))]$, we obtain

$$\lambda(x) = \delta(x) + \frac{E[R/N|X = x]}{E\{1/[\theta C''(y(x, \theta))]|X = x\}}. \tag{4.11}$$

Because C'' is not identified by our previous result, $\lambda(x)$ is not identified nonparametrically. On the other hand, $\lambda(x)$ is identified as soon as one imposes parametric restrictions on C' and $F_{\theta|X,S}$. The following proposition shows that in turn, all parameters defined above are identified.

PROPOSITION 4.1. *Suppose that C' is point identified on $(0, 1)$. Then $\lambda(x)$, $\Pi(x)$, the counterfactual transfers defined in (4.6), (4.8), and (4.10), and the corresponding surpluses are identified.*

4.3 *Parametric estimation of the cost function and the distribution of types*

We thus consider a parametric estimation of $F_{\theta|X,S}$ and C'. An important aspect is that we use the nonparametric estimates $(C'(y_k), F_{\theta|X,S}(\theta_k))_{k \in \mathcal{K}}$ to investigate which parametric family fits best. We compare three standard family of distributions on $\mathbb{R}^+$ for $F_{\theta|X,S}$, namely the Fréchet, for which $F_{\theta|X,S}(\theta) = \exp(-a\theta^{-b})$ $(a, b > 0)$, the log-normal, for which $F_{\theta|X,S}(\theta) = \Phi((\ln \theta - a)/b)$ (where Φ denotes the cumulative distribution function of a standard normal variable and $b > 0$) and the Weibull, for which $F_{\theta|X,S}(\theta) = 1 - \exp(-a\theta^b)$ $(a, b > 0)$.[23] These families differ in their tail behavior. The first has heavy tails (power ones), the second medium tails (between power and exponential ones), and the third light tails (exponential ones).

---

[23]We also tried compactly supported (Beta) distributions, but their fit was less satisfactory.
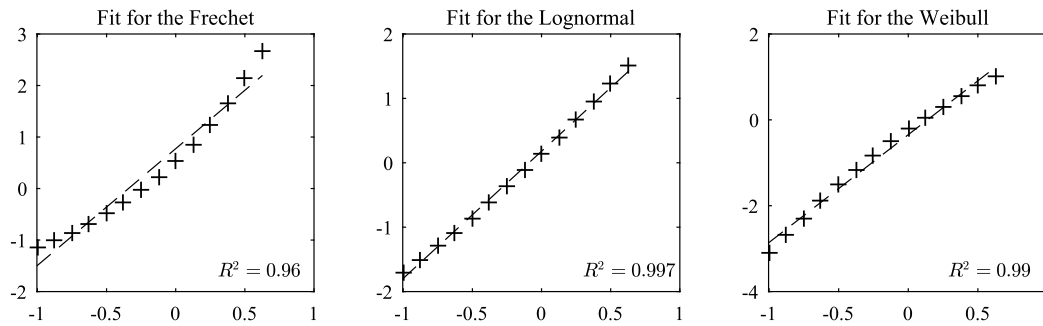
FIGURE 5. Choice of the parametric family for $F_{\theta|S=1}$. *Notes*: plot of $-\ln[-\ln \widehat{F}_{\theta|S=1}(\theta_k)]$ (left graph), $\Phi^{-1}(\widehat{F}_{\theta|S=1}(\theta_k))$ (middle graph), and $\ln[-\ln(1-\widehat{F}_{\theta|S=1}(\theta_k))]$ (right graph) against $\ln(\theta_k)$. Points should be aligned for the true parametric family. The dotted lines are the best linear approximations.

To choose between these three parametric families, we plot $-\ln(-\ln F_{\theta|X,S}(\theta_k))$, $\Phi^{-1}(F_{\theta|X,S}(\theta_k))$, and $\ln(-\ln(1-F_{\theta|X,S}(\theta_k)))$ against $\ln \theta_k$. The points should be aligned if the parametric family is the true one. For instance, in the case of a Fréchet distribution, $-\ln(-\ln F_{\theta|X,S}(\theta_k)) = \ln(a) - b \ln \theta_k$. Similarly, we consider families of marginal cost functions tending to 0 at 0 and to ∞ at 1, but which differ in their behavior at infinity. We consider $C'(y) = \alpha \phi(y/(1-y))^\beta$, with $\phi(x) = \ln(1+x)$, x or $\exp(x) - 1$. Once more, we plot $\ln C'(y_k)$ against $\ln \phi(y/(1-y))$ in the three cases. The true function ϕ should satisfy $\ln C'(y_k) = \ln \alpha + \beta \ln \phi(y_k/(1-y_k))$. Figures 5 and 6 display the three corresponding plots. They indicate that the lognormal distribution and $\phi(x) = \ln(1+x)$ have the best fits. The lognormal distribution is also preferred for the movers ($X = 1, S = 0$ and $X = 2, S = 0$).

With these parametric specifications at hand, we can estimate the model by maximum likelihood, assuming also that stayers have the same θ at both periods. To take into account possible differences between the stayers and the others, we suppose that $\ln \theta | X = x, S = 0 \sim \mathcal{N}(a_x, b)$, while $\ln \theta | S = 1 \sim \mathcal{N}(a_3, b)$. We denote by $\eta =$
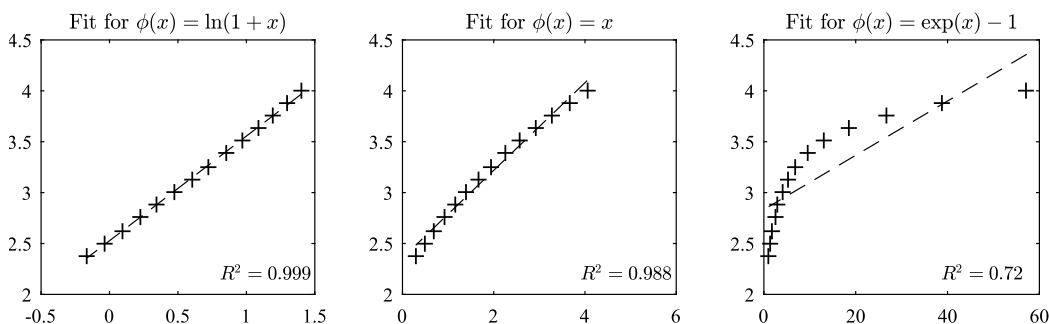

FIGURE 6. Choice of the parametric family for C'. *Notes*: plots of $\ln \widehat{C}'(y_k)$ against $\ln \phi(y_k/(1-y_k))$ for different choices of ϕ. The points should be aligned for the true function ϕ. The dotted lines are the best linear approximations.

$(\alpha, \beta, a_1, a_2, a_3, b)$ the vector of parameters to estimate. To compute the maximum like-lihood estimator, note first that

$$y(x, \theta|\eta) = 1 - \exp\left(-\left(\frac{\delta(x)}{\alpha\theta}\right)^{1/\beta}\right). \tag{4.12}$$

Let R_x and N_x denote the number of respondents and the number of households to interview for survey x. The likelihood of observing $(R_1 = r_1, R_2 = r_2)$ conditional on (N_1, N_2) for a stayer satisfies

$$\Pr(R_1 = r_1, R_2 = r_2|N_1, N_2, S = 1, \eta)$$

$$= \binom{r_1}{N_1}\binom{r_1}{N_1} E\big[y(1, \theta|\eta)^{r_1}(1 - y(1, \theta|\eta))^{N_1-r_1}$$

$$\times y(2, \theta|\eta)^{r_2}(1 - y(2, \theta|\eta))^{N_2-r_2}|S = 1\big]$$

$$= \binom{r_1}{N_1}\binom{r_2}{N_2} \int y(1, \theta|\eta)^{r_1}(1 - y(1, \theta|\eta))^{N_1-r_1}$$

$$\times y(2, \theta|\eta)^{r_2}(1 - y(2, \theta|\eta))^{N_2-r_2} f_{\theta|S=1}(\theta|\eta)\, d\theta.$$

The likelihood for movers write similarly, except that it only involves (R_1, N_1) or (R_2, N_2). Because the log-likelihood includes integrals that do not have closed forms in general, we use simulations to approximate it. Once we obtain $\widehat{\eta}$, we use it to derive an estimator of $\lambda(x)$ and of all policy parameters, using plug-in estimators and the formulas above.

The maximum likelihood estimates of η under the parameter specification chosen above are displayed in Table 7. We first estimate the model without constraint on (a_1, a_2). The first column of Table 7 shows that $\widehat{a}_3$ is smaller than $\widehat{a}_1$ and $\widehat{a}_2$, reflecting an average better productivity of the stayers. $\widehat{a}_1$ and $\widehat{a}_2$ are close to each other and not statistically different (p-value $= 0.34$). This result is in line with the one of the nonparametric test of no selection. To obtain more accurate results, we then reestimate the model under the constraint that $a_1 = a_2$ (column 2 of Table 7). The results are very similar. Under this latter specification, we obtain $\widehat{\lambda}_1 = 89.3$ and $\widehat{\lambda}_2 = 110.8$, the higher value of $\widehat{\lambda}_2$ reflecting the higher importance for Insee of the 2003 survey.

4.4 *The cost of using inefficient contracts*

We now turn to the results on surpluses, computed using the constrained estimates above. We focus on the 2003 survey, the results being very similar for 2001–2002. Table 8 summarizes our results. First, we find that the surplus loss associated with the use of linear contracts is around 16% (68.3 versus 80.9) and that the response rate decreases by 10% compared to optimal contracts (83% versus 93%). This result contrasts with the idea that simple contracts can be quite inefficient. Ferrall and Shearer (1999), for instance, evaluated 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

TABLE 7. Maximum likelihood estimates of the parameters of $C'(y) = \alpha[\ln(1 + y/(1 - y))]^\beta$ and $F_{\theta|X=x,S=s}(\theta) = \Phi(\frac{\ln(\theta)-a_{1+(1-s)x+2s}}{b})$.

Parameter	Unconstrained Estimates	Constrained Estimates
a_1	0.022 (0.053)	0.053 (0.045)
a_2	0.11 (0.07)	same as a_1
a_3	−0.06 (0.04)	−0.07 (0.05)
b	0.44 (0.11)	0.48 (0.12)
α	11.66 (1.60)	11.14 (1.68)
β	1.16 (0.29)	1.25 (0.32)
λ_1	84.16 (16.69)	89.31 (18.46)
λ_2	103.79 (19.35)	110.80 (21.25)

Note: 614 observations. The constrained model rules out any selection effects by imposing $a_1 = a_2$. The standard errors are computed using the inverse of the estimated Hessian, and the delta method.

highly nonlinear. This may explain why firms widely use linear contracts compared to nonlinear ones: they are less costly to implement and almost efficient. This result is in line with a result of Miravete (2007), who reports a loss of only 3%, also supports this claim. It is also consistent 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 proved that simple English auctions generate an expected price that is more than 80% of the value of the object to the bidder with the highest valuation. Finally, studying mixed bundling, Chu, Leslie, and Sorensen (2011) showed that simple pricing strategies are often nearly optimal. With surprisingly few prices, a firm can obtain 99% of the profit that would be earned by mixed bundling.

   In line with these results, we find that Insee can use simple contracts and still give the right incentives to its interviewers. Also, a change in the rewarding scheme could have been difficult to implement for Insee. Our results imply indeed that only 4.2% of

TABLE 8. Surplus and response rates under alternative compensation schemes.

Environment	Pay Method	E[surplus]	Relative	E[response rate]
Full information	Optimal contract	103.04 (21.18)	1.00 (0)	0.99 (0.005)
Incomplete information, no moral hazard	Optimal contract	80.86 (18.89)	0.78 (0.02)	0.93 (0.01)
Incomplete information	Optimal contract	80.86 (18.89)	0.78 (0.02)	0.93 (0.01)
Incomplete information	Linear contract	68.30 (17.37)	0.66 (0.03)	0.83 (0.007)

Note: 614 observations. The standard errors are computed using the delta method.
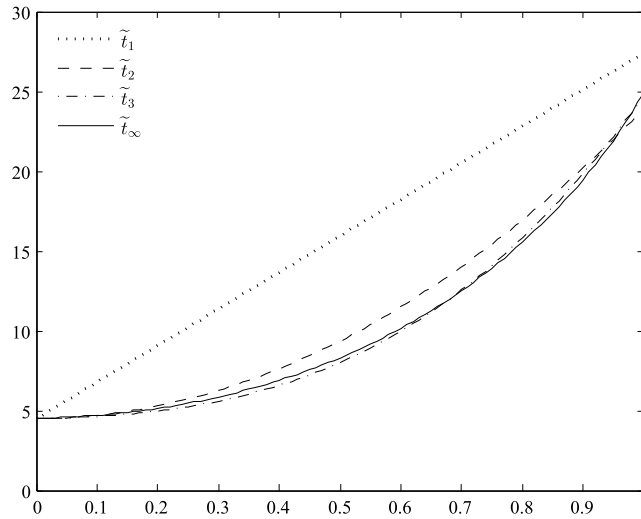
FIGURE 7. Evolution of $\widetilde{t}_n = E[t_{r_2}/n_2 | n_2 = n, y]$ with n.

the interviewers (the less efficient ones) would have benefitted from a change from the linear contract used in 2003 to the optimal, nonlinear contract.

Second, we find no significant cost of moral hazard here. If Insee was able to use contracts based on the probability of response rather than on the realized number of respondents, it would only increase its surplus by 0.0037% only. To understand this, recall that the cost of moral hazard is small when absent any moral hazard, the optimal polynomial contracts

$$\widetilde{t}_n(y) \equiv \frac{1}{n} \sum_{k=0}^{n} \binom{n}{k} y_n(x, \theta, t)^k \big(1 - y_n(x, \theta, t)\big)^{n-k} t_{nk}^*,$$

where $t_n^* = (t_{n0}^*, \ldots, t_{nn}^*)$, can approximate well the unconstrained optimal contracts $\widetilde{t}_\infty(y) \equiv t_n^{WM}(x, y)/n$.[24] In Figure 7, we plot the functions $\widetilde{t}_n$ for $n = 1, 2, 3$, and $n = +\infty$. While there is an important gap between $n = 1$ and $n = 2$, $\widetilde{t}_2$ provides already a good approximation of $\widetilde{t}_\infty$, while the fit is almost perfect for $n = 3$. This explains the overall negligible loss, as the number of households for which $n \leq 2$ only represent 0.14% of the whole sample of households.

Third, we find moderate cost of incomplete information, the optimal surplus under asymmetric information being 78% of the optimal one under full information. This loss of 22% is in particular smaller than the one reported by Ferrall and Shearer (33%). Moreover, the surplus under asymmetric information and with the linear contract is 66% of what it could be under complete information. The main part of this loss (65%) is due to incomplete information whereas 35% is associated with the simple tarification.

The rather mild degree of asymmetric information between Insee and its interviewers may explain why Insee chooses not to use some information at its disposal. To confirm this intuition, we investigate what Insee would obtained if it relied on interviewers'

---

[24]One can show using (4.7) and (4.8) that $t_n^{WM}(x, y)/n$ does not depend on n.

TABLE 9. Estimation of the parametric model with interviewers' co-variates.

Variable	Unrestricted Model	Restricted Model
Intercept	−0.005 (0.093)	−0.102 (0.065)
Experience ≤ 5	0.333 (0.103)	0.333 (0.097)
Experience between 5 and 15	0.194 (0.077)	0.186 (0.071)
Rural or small urban area	−0.229 (0.074)	−0.212 (0.066)
Female	0.001 (0.066)	–
Married	−0.053 (0.051)	–
Stayer	−0.089 (0.059)	–
σ	0.421 (0.107)	0.398 (0.096)
β	1.222 (0.304)	1.156 (0.27)
$R^2 = 1 - \sigma^2 / V(\ln \theta)$	0.158	0.145

Note: 600 and 601 observations are used in the unrestricted and restricted model. A positive parameter indicates larger θ, and thus, on average, lower response rates. The standard errors are computed using the inverse of the estimated hessian, and the delta method.

characteristics. We first estimate how the characteristics W of an interviewer relate to θ, by positing

$$\ln \theta = W\gamma + \nu,$$

where, in line with our lognormal specification, we suppose $\nu | W \sim \mathcal{N}(0, \sigma^2)$. The characteristics include experience, gender, marriage status, the interviewer's status, and the type of area (large urban areas versus others).[25] We reestimate the model keeping our preferred specification of the cost function. The results are displayed in Table 9. Not surprisingly, we find that interviewers with larger experience and living in smaller urban areas perform better on average. Gender and marital status do not seem to be correlated with interviewers' types. Once controlling for experience and the type of area, we also see that stayers do not perform better on average. This can be seen as a confirmation of our previous result that participation's decision is not endogenous. Overall, the part of the variance of interviewers' type that is explained by their observable characteristics is quite small, around 15%. Note also that the estimators of β is very similar to the one we obtained before.

These results suggest that experience and the type of areas are the major determinants of interviewers's type. Using the same model restricted to these covariates (see the second column of Table 9), we estimate what would be the optimal bonus to provide to the six types of interviewers defined by the interactions of these two variables. Insee would propose bonuses ranging from 20.2 for interviewers with more than 15 years

[25]We do not include the dummy of having another professional activity because of missing data. The interviewer's area is considered as a large urban areas if most of the housings he has to interview are in towns with more than 100,000 inhabitants.

of experience in rural or small urban areas to 25.5 for interviewers working for Insee for less than 5 years in large urban areas. Overall, however, the gain in terms of surplus would remain nearly constant, with a negligible gain of only 0.15%. This very small gain can be explained by two things. First, the characteristics we use only explain 6.3% of the variance of the interviewers' types. The adverse selection problem remains therefore relatively important. Second, we still consider linear contracts here, and they are not optimal. At the end, the cost of discriminating between interviewers is thus likely to exceed these expected gains. In addition to implementation costs mentioned by Ferrall and Shearer (1999), Insee faces social costs due to quite strong unions opposed to such discriminations.

5. Conclusion

This work contributes to the empirical personnel literature by showing, in a context of moderate asymmetric information, that interviewers react to incentives and that the simple contracts proposed by Insee are nearly optimal. Beyond these empirical results, we also propose a new approach that extensively uses the exogenous change in 2003 in the compensation scheme, the piece rate increasing from 20.2 to 22.9 euros. This change allows us, in particular, to identify and recover nonparametrically some information on the cost function of the interviewers and on the distribution of their types. This information is used to select correctly the parametric restrictions that we need to impose to derive our results. More generally, we believe that such an exogenous change, associated with a nonparametric estimation in a first step, is essential to estimate and test the optimality of contracts or the presence of asymmetries.

Appendix A: Details on interviewers data

Besides the data on the surveys, we also have some limited data on interviewers who participate to these surveys and, more generally, on Insee's households interviewers at the beginning 2001. The most striking fact emerging from Table 10 is the large average experience of interviewers: 8.5 years for the whole set of interviewers and around 10 years for PCV interviewers. Moreover, out of the 12 surveys conducted by Insee in 2001 and for which we have information about interviewers, a typical interviewer conducts more than 5 surveys a year in his designated area. This is not surprising, given that Insee basically relies on the same pool of interviewers for all its surveys, even if the precise set of interviewers may vary from one survey to another. By doing so, Insee avoids sunk costs stemming from the recruitment of new interviewers. This sunk cost includes the recruitment procedure itself, as well as a 3-day training period received by interviewers before they conduct their first survey. A second reason is that experience matters for this job. It is well documented that interviewers may influence households and bias their responses (see, e.g., Mensh and Kandel (1988) or O'Muircheartaigh and Campanelli (1998)). It seems, however, that experienced interviewers are less prone to this so-called interviewer's effect (see, e.g., Cleary, Mechanic, and Weiss (1981), Singer, Frankel, and Glassman (1983), or Campanelli, Martin, and Rothgeb (1991)). Finally, most surveys

TABLE 10. Descriptive statistics on Insee household interviewers.

	All Interviewers in 2001	PCV Interviewers		
Variable		2001	2002	2003
Experience at Insee (in years)	8.55 (6.97)	10.47 (6.52)	9.7 (6.94)	9.56 (7.24)
Yearly income	4054 (3075)	6146 (3139)	2976 (1702)	5169 (2300)
Number of surveys done during the year	5.45 (3.73)	8.59 (2.89)	8.21 (3.85)	7.97 (2.82)
Female	0.84 (0.37)	0.85 (0.35)	0.85 (0.35)	0.84 (0.36)
Married	0.66 (0.47)	0.69 (0.46)	0.69 (0.46)	0.68 (0.47)
Age	47.9 (9.07)	49.7 (7.86)	49.2 (8.00)	49.5 (8.49)
Other professional activity	0.41 (0.49)	0.39 (0.49)	0.40 (0.49)	0.42 (0.49)
Number of obs.	939	379	469	453

Note: For each column, we indicate the mean and standard deviation (in parenthesis) of the variables. Some observations are missing for the dummy of other professional activity. The income is computed using most but not all of the household surveys.

are repeated over time. As interviewers receive a specific training corresponding to each survey, relying on the same pool of interviewers from one edition to another also allows Insee to avoid the duplication of these training costs.

Table 10 also shows that the typical interviewer is a middle-aged woman who is out of the labor market. Conversations with them reveal that their job at Insee is usually not the main source of income for the household. It is a flexible job that allows them to complement the revenue of the family. Even if there is a large variability among interviewers and across years, the annual income of 4545 euros earned on average by household interviewers in 2001 corresponds to the minimum wage for a third time job.

Appendix B: The effects of the interviewers' sample size n

B.1 *Alternative test of Assumption* 3

We consider another test of (3.5), based not only on the conditional expectation of R, but on its whole distribution. We rely for that purpose also on the binomial model posited in Assumption 1.[26] These two conditions imply a known link between $\Pr(R = r|N = n, X = x, S = s)$ and the first n moments of $y(x, \theta)$. Specifically, integrating equation (C.1) below over S leads to

$$\mathbf{P}_n^x = \mathbf{Q}_n m_n^x,$$ (B.1)

where $\mathbf{P}_n^x = (\Pr(R = 1|X = x, N = n), \ldots, \Pr(R = n|X = x, N = n))'$, $\mathbf{Q}_n$ is a nonsingular matrix whose terms are given in the proof of Theorem 3.1 below and $m_n^x = (E(y(x, \theta)^1|X = x), \ldots, E(y(x, \theta)^n|X = x))'$. Equation (B.1) should hold for all $n \in$

---

[26]The previous test only uses the first moment of R. So it is not relying on the independence in households reactions, but just on the fact any household within an area had the same probability of being interviewed.

$\mathrm{Supp}(N|X = x)$. Because m_n^x and $m_{n'}^x$ have $n \wedge n'$ common terms, many overidentifying restrictions are available. We thus consider a test close to usual overidentification tests for minimum distance estimation. A difference, though is that we also incorporate the constraint that m_n^x should be a vector of moments, which implies several restrictions such as variance positivity. We refer to D'Haultfœuille and Rathelot (2017) for details on how to incorporate these constraints.[27] If some of these constraints are binding, the test statistic does not have an asymptotic chi-squared distribution. To estimate the critical value, we therefore draw bootstrap samples under the null distribution. D'Haultfœuille and Rathelot (2017) established the validity of a very similar bootstrap test (see their Theorem C.1). At the end, we obtain p-values of 0.93 and 0.71 for the two surveys, supporting again the validity of Assumption 3.

B.2 *Partial identification of marginal costs without Assumption* 3

We show here that we can actually weaken the condition that $C(n, x, y)/C(x, y) = n$ and still obtain bounds on the effect of n on the cost function. Specifically, let us assume that $C(n, x, y) = f(n)C(x, y)$ for some function $f(\cdot)$ satisfying without loss of generality $f(n_0) = 1$ for some $n_0 \in \mathrm{Supp}(N|X = 1) \cap \mathrm{Supp}(N|X = 2)$. Combining this with our exclusion restriction $C(x, y) = C(y)$, we obtain

$$y(n, x, \theta) = C'^{-1}\left(\frac{n\delta(x)}{f(n)\theta}\right).$$

Then, for all $x \in \{1, 2\}$ and $n \in \mathrm{Supp}(N|X = x)$,

$$\frac{E(R|N = n, X = x)}{n} = E\left[C'^{-1}\left(\frac{n\delta(x)}{f(n)\theta}\right)\right].$$

Moreover, C'^{-1} is strictly increasing. This means that for any $(n_1, n_2) \in \mathrm{Supp}(N|X = 1) \times \mathrm{Supp}(N|X = 2)$,

$$\mathrm{sgn}\left[\frac{E(R|N = n_1, X = x)}{n_1} - \frac{E(R|N = n_2, X = x)}{n_2}\right] = \mathrm{sgn}\left(\frac{n_1\delta(1)}{f(n_1)} - \frac{n_2\delta(2)}{f(n_2)}\right).$$

This allows us to obtain bounds on $f(\cdot)$, and possibly point identify $f(\cdot)$ if this function is parametrized.

B.3 *The effect of a bounded support on* N

The point identification in Theorem 3.1 relies on a large support assumption on N (see Assumption 4). If $\sup \mathrm{Supp}(N|X = x) = \overline{N}_x < +\infty$, the proof of Theorem 3.1 reveals that we identify only the first $\overline{N}_X$ moments of $Y|X, S$. This implies that the distribution of

---

[27]Also, we actually perform our test conditional on $N < 35$. We faced numerical issues otherwise, due in particular to constrained optimization in large dimensional spaces. This restriction only removes less than 2% of the data in the two cases.
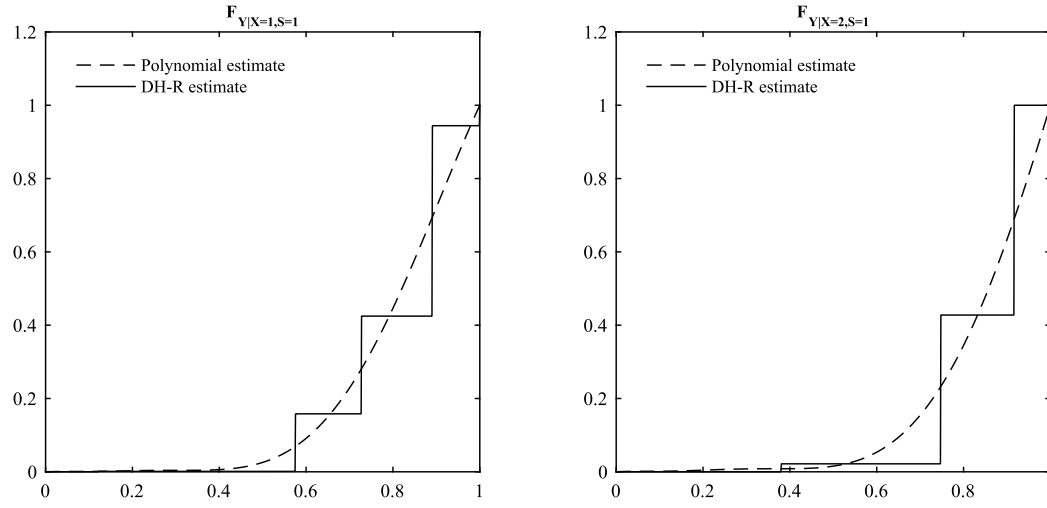
FIGURE 8. D'Haultfœuille and Rathelot (DH-R)'s and sieve ML estimates of $F_{Y|X=x,S=1}$.

$Y|X, S$ is not point identified in general. We can still obtain bounds on $F_{Y|X,S}(y)$ by minimizing or maximizing $F(y)$ over all cumulative distributions F with prescribed first $\overline{N}_X$ moments, taking also into account that $F_{Y|X=2,S=1}(y) \leq F_{Y|X=1,S=1}(y)$ for all y.

   This problem is difficult and, to our knowledge, has not been addressed in the literature. On the other hand, if we do not include the inequality constraints, we can use Theorem 2.1 of D'Haultfœuille and Rathelot (2017), who show that the optimization can be conducted without loss of generality on discrete distributions with at most $\overline{N}_X + 1$ support points. Of course, the corresponding bounds $\underline{F}_{Y|X,S}(y)$ and $\overline{F}_{Y|X,S}(y)$ are not sharp, as they do not incorporate the inequality constraints above.

   We computed the estimators of $\underline{F}_{Y|X,S}(y)$ and $\overline{F}_{Y|X,S}(y)$ suggested by D'Haultfœuille and Rathelot (2017). The results are displayed in Figure 8, which also presents for comparison the sieve estimates considered in Section 3.4. We actually obtain a point estimate on $F_{Y|X,S}$. This could be expected, given that here $\overline{N}_X \geq 50$. D'Haultfœuille and Rathelot (2017) showed in their setting that the upper and lower estimated bounds typically collapse for usual sample sizes (up to 10,000, say) when $\overline{N}_X \geq 6$. Also, the estimator corresponds to that of a finitely supported distribution, which is also expected given that optimization is run over distributions with at most $\overline{N}_X + 1$ support points. Other than this feature, the estimator looks quite similar to the sieve estimator.

APPENDIX C: PROOFS

PROOF OF THEOREM 3.1. First, by Assumption 3 and the first-order condition, $y(x, \theta)$ does not depend on n. We denote it by $y(x, \theta)$. Then, for all n in the support of $N|X = x, S = s$ and all $1 \leq r \leq n$, we have

$$\Pr(R = r|N = n, X = x, S = s) = E\big[\Pr\big(R = r|N = n, y(x, \theta)\big)|N = n, X = x, S = s\big]$$

$$= E\left[\binom{n}{r} y(x, \theta)^r \left(1 - y(x, \theta)\right)^{n-r} | N = n, X = x, S = s\right]$$

$$= E\left[\binom{n}{r} y(x, \theta)^r \left(1 - y(x, \theta)\right)^{n-r} | X = x, S = s\right]$$

$$= \sum_{i'=0}^{n-r} \binom{n}{r} \binom{n-r}{i'} (-1)^{n-r-i'} E\left(y(x, \theta)^{n-i'} | X = x, S = s\right)$$

$$= \sum_{i=r}^{n} \binom{n}{i} \binom{i}{r} (-1)^{i-r} E\left(y(x, \theta)^i | X = x, S = s\right)$$

$$= \sum_{i=1}^{n} \binom{n}{i} \binom{i}{r} (-1)^{i-r} E\left(y(x, \theta)^i | X = x, S = s\right),$$

where the first equality follows from the law of iterated expectation, the second from Assumption 1, the third stems from independence between θ and N conditional on $X = x, S = s$ (Assumption 3), the fourth from the decomposition of $(1 - y(x, \theta))^{n-r}$, the fifth is obtained by setting $i = n - i'$ and remarking that $\binom{n}{r}\binom{n-r}{i'} = \binom{n}{i}\binom{i}{r}$, and the last by noting that the $j - 1$ first terms in the sum are zero. Hence, letting $\mathbf{P}_n^{x,s} = (\Pr(R = 1 | N = n, X = x, S = s), \ldots, \Pr(R = n | N = n, X = x, S = s))'$, $m_n^{x,s} = (E(y(x, \theta)^1 | X = x, S = s), \ldots, E(y(x, \theta)^n) | X = x, S = s)'$ and $\mathbf{Q}_n$ be the $n \times n$ matrix of typical (i, r) element $\binom{n}{i}\binom{i}{r}(-1)^{i-r}$, we get

$$\mathbf{P}_n^{x,s} = \mathbf{Q}_n m_n^{x,s}. \tag{C.1}$$

Moreover, $\mathbf{Q}_n$ is invertible as an upper triangular matrix with nonzero diagonal elements. Thus, $m_n^{x,s}$ is identified by $\mathbf{Q}_n^{-1}\mathbf{P}_n^{x,s}$. Conditional on $X = x, S = s$, the n first moments of $y(x, \theta)$ are identified from the distribution of R conditional on $N = n, X = x, S = s$. Because $\sup\{n : \Pr(N = n | X = x, S = s) > 0\} = +\infty$, all moments of $y(x, \theta)$ (conditional on $X = x, S = s$) are identified. This, together with $y(x, \theta)$ bounded, ensures that the distribution of $y(x, \theta)$ conditional on $X = x, S = s$ is identified (see, e.g., Gut (2005)). □

PROOF OF THEOREM 3.2. It follows from the discussion before Theorem 3.2 that C', $F_{\theta|X=1,S=s}$ and $F_{\theta|X=2,S=s}$ are point identified on $(y_k)_{k \in \mathcal{K}}$, $(\theta_k)_{k \in \mathcal{K}}$ and $(\theta_k)_{k+1 \in \mathcal{K}}$, respectively. Elsewhere,

$$\theta_{\overline{k}(y)} = \sup_{k \in \mathcal{K}: y_k \geq y} \theta(1, y_k) \leq \theta(1, y) \leq \inf_{k \in \mathcal{K}: y_k \leq y} \theta(1, y_k) = \theta_{\underline{k}(y)}.$$

Thus,

$$C'(y) = \delta(1)/\theta(1, y) \geq \delta(1)/\theta_{\underline{k}(y)} = \underline{C}'(y),$$

and similarly, $C'(y) \leq \overline{C}'(y)$. Besides,

$$y_{\underline{k}(\theta)} = \sup_{k \in \mathcal{K}: \theta_k \geq \theta} y_k \leq y(1, \theta) \leq \inf_{k \in \mathcal{K}: \theta_k \leq \theta} y_k = y_{\overline{k}(\theta)}.$$

Hence,

$$F_{\theta|X=1,S=s}(\theta) = 1 - F_{Y|X=1,S=s}\big(y(1,\theta)\big) \geq 1 - F_{Y|X=1,S=s}(y_{\overline{k}(\theta)}) = \underline{F}_{\theta|X=1,S=s}(\theta),$$

and similarly for the upper bound. The bounds on $F_{\theta|X=2,S=s}(\theta)$ follow by remarking that

$$y(2,\theta) = y\big(1,\delta(1)\theta/\delta(2)\big) \in [y_{\underline{k}(\theta)+1}, y_{\overline{k}(\theta)+1}].$$

We now show that for all $y^0 \in (0,1)\backslash\{y_k : k \in \mathcal{K}\}$, and $\theta^0 \in \mathbb{R}^+\backslash\{0, \theta(1,y_k) : k \in \mathcal{K}\}$, the bounds on $C'(y^0)$ and $F_{\theta|X,S}(\theta^0)$ are sharp. We focus on $\overline{C}'(y^0)$ as the proof is similar for $\underline{C}'(y^0)$, $\overline{F}_{\theta|X,S}(\theta^0)$ and $\underline{F}_{\theta|X,S}(\theta^0)$. More precisely, we want to construct a function $\widetilde{C}'$ such that $\widetilde{C}'(y^0)$ is arbitrarily close to $\overline{C}'(y^0)$ and all the restrictions given by the data and the model hold. We consider separately two cases, whether or not there exists $k \in \mathcal{K}$ such that $y_k < y^0 < y_{k+1}$.

In the first case, fix ε such that $0 < \varepsilon < \delta(1)[1/\theta_{k+1} - 1/\theta_k]$. We first define $\widetilde{C}'$ on $[y_k, y_{k+1})$. To do so, we consider any strictly increasing, continuously differentiable function $\widetilde{C}'$ such that $\widetilde{C}'(y_k) = \delta(1)/\theta_k$, $\widetilde{C}'(y^0) = \overline{C}'(y^0) - \varepsilon$, $\lim_{y\uparrow y_{k+1}} \widetilde{C}'(y) = \delta(1)/\theta_{k+1}$, and

$$\lim_{y\uparrow y_{k+1}} \widetilde{C}''(y) = \frac{\delta(2)\widetilde{C}''(y_k)}{\delta(1)H'(y_{k+1})}. \tag{C.2}$$

Such a function exists because $\delta(1)/\theta_k < \overline{C}'(y^0) - \varepsilon < \delta(1)/\theta_{k+1}$ and $H(y) = C'^{-1}[\delta(2) \times C'(y)/\delta(1)]$ is differentiable with positive derivative at any $y > 0$.

We then extend $\widetilde{C}'$ on $(0,1)$ using (3.3). For instance, assume that $k + 2 \in \mathcal{K}$. Then we define $\widetilde{C}'$ on $[y_{k+1}, y_{k+2})$ by

$$\widetilde{C}'(y) = \frac{\delta(2)}{\delta(1)}\widetilde{C}'\big(H^{-1}(y)\big).$$

Moreover, because H is continuously differentiable, $\widetilde{C}'$ is continuously differentiable on (y_{k+1}, y_{k+2}). It also admits a right derivative at y_{k+1} given by

$$\lim_{y\downarrow y_{k+1}} \widetilde{C}''(y) = \frac{\delta(2)\widetilde{C}''(y_k)}{\delta(1)H'(y_k)},$$

and equation (C.2) ensures that $\widetilde{C}'$ is differentiable at y_{k+1}. By induction, using either H or H^{-1}, we can then extend $\widetilde{C}'$ on $\mathcal{Y}_\mathcal{K} \equiv \bigcup_{k:k\in\mathcal{K},k+1\in\mathcal{K}}[y_k, y_{k+1})$. If $\mathcal{Y}_\mathcal{K} = (0,1)$, we have defined this way a continuously differentiable function on the whole interval $(0,1)$. This function is also strictly increasing as both H or H^{-1} are strictly increasing. If $\mathcal{Y}_\mathcal{K} \neq (0,1)$, we still have to extend $\widetilde{C}'$ on intervals of the form $[0, \underline{y})$ or $[\overline{y}, 1)$. Consider the first case (the second is similar). We simply consider any strictly increasing, continuously differentiable function $\widetilde{C}'$ such that $\widetilde{C}'(0) = 0$, $\lim_{y\uparrow \underline{y}} \widetilde{C}'(y) = \widetilde{C}'(\underline{y})$, and $\lim_{y\uparrow \underline{y}} \widetilde{C}''(y) = \widetilde{C}''(\underline{y})$. Again, this defines a continuously differentiable function on the whole interval $(0,1)$.

To complete the proof for this case, we have to show that with such a function $\widetilde{C}'$, we can rationalize the model and the data. For that purpose, let $\widetilde{\theta}(x,y)$ be defined

by $\widetilde{\theta}(x, y) = \delta(x)/\widetilde{C}'(y)$ for all $y \in \mathcal{Y}_x$. By construction, $\widetilde{\theta}(x, \cdot)$ is strictly decreasing. Let $\widetilde{y}(x, \cdot)$ denote its inverse, and let

$$\widetilde{F}_{\theta|X=x,S}(\theta) = 1 - F_{Y|X=x,S}\big(\widetilde{y}(x, \theta)\big).$$

By construction, $\widetilde{C}'$ and $\widetilde{F}_{\theta|X,S}$ rationalize the data and the first-order condition (3.2). The second-order condition also holds since $\widetilde{C}'$ is strictly increasing. Thus, these functions rationalize Model (3.1) as well. Because ε could be arbitrarily close to 0, this shows that $\overline{C}'(y)$ is sharp.

We now consider the case where there is no $k \in \mathcal{K}$ such that $y_k < y^0 < y_{k+1}$. Equivalently, $y^0 \notin \mathcal{Y}_\mathcal{K}$, and either $y^0 \in [0, \underline{y})$ or $y^0 \in (\overline{y}, 1)$. In both case, we simply let $\widetilde{C}' = C'$ on $\mathcal{Y}_\mathcal{K}$. Suppose that $y^0 \in [0, \underline{y})$ (the case $y^0 \in (\overline{y}, 1)$ is similar). Note that $\underline{y} \in (y_k)_{k \in \mathcal{K}}$, which implies that $\overline{C}'(y^0) = C'(\underline{y})$. Fix ε such that $0 < \varepsilon < C'(\underline{y})$. Then define $\widetilde{C}'$ on $[0, \underline{y})$ as any strictly increasing, continuously differentiable function such that $\widetilde{C}'(0) = 0$, $\widetilde{C}'(y^0) = \overline{C}'(y^0) - \varepsilon$, $\lim_{y\uparrow\underline{y}} \widetilde{C}'(y) = C'(\underline{y})$ and

$$\lim_{y\uparrow\underline{y}} \widetilde{C}''(y) = C''(\underline{y}).$$

In the case where $\sup \mathcal{Y}_\mathcal{K} < 1$, define similarly $\widetilde{C}'$ on $[\overline{y}, 1)$ by $\widetilde{C}'(\overline{y}) = C'(\overline{y})$, $\lim_{y\uparrow 1} \widetilde{C}'(y) = +\infty$ and $\lim_{y\downarrow\overline{y}} \widetilde{C}''(y) = C''(\overline{y})$. By construction, $\widetilde{C}''$ is then strictly increasing and continuously differentiable on $(0, 1)$. The rest of the proof is identical as above. $\quad\square$

NONIDENTIFICATION WITH ONE MENU OF CONTRACTS. With only one menu of contracts, the variables X and S are irrelevant, so we drop them here. Let us consider a strictly increasing and differentiable function $\widetilde{C}'$, different from the true one C'. Define then $\widetilde{\theta}(y)$ by $\widetilde{\theta}(y) = \delta/\widetilde{C}'(y)$. $\widetilde{\theta}$ is strictly decreasing and admits an inverse function $\widetilde{y}$. Then define $\widetilde{F}_\theta$ by

$$\widetilde{F}_\theta(\theta) = 1 - F_Y\big(\widetilde{y}(\theta)\big).$$

By construction, $\widetilde{C}'$ and $\widetilde{F}_\theta$ are consistent with the first- and second-order conditions and the identified distribution F_Y. As a result, C' and $\widetilde{F}_\theta$ are not identified. $\quad\square$

PROOF OF THEOREM 3.3. The proof proceeds in four steps. We first prove that $\widehat{F}_{Y|X=x,S=s}$ is uniformly consistent. We then prove that $\widehat{H}$ is uniformly consistent on each compact set included in $(0, 1)$. Third, we prove that for all $k \in \mathcal{K}$, $\widehat{y}_k$ is consistent. Finally, we show that the estimated bounds of C' and $F_{\theta|X,S}$ are consistent.

1. Uniform consistency of $\widehat{F}_{Y|X=x,S=s}$.

For any function g on $[0, 1]$ let $\|g\| = \sup_{x \in [0,1]} |g(x)|$. We actually prove the stronger result that for all $(x, s) \in \{1, 2\} \times \{0, 1\}$,

$$\|\widehat{f}_{Y|X=x,S=s} - f_{Y|X=x,S=s}\| \xrightarrow{\mathbb{P}} 0. \tag{C.3}$$

For all y in the interior of $\mathcal{Y}_{x,s}$, $f_{Y|X=x,S=s}(y) = \partial\theta/\partial y(x,y)f_{\theta|X=x,S=s}(\theta(x,y))$. Hence, $f_{Y|X=x,S=s}$ is continuous in the interior of $\mathcal{Y}_{x,s}$. Moreover, differentiating the first-order condition, we obtain

$$\frac{\partial\theta}{\partial y}(x,y) = -\frac{\theta(x,y)C''(y)}{C'(y)} = -\frac{\delta(x)C''(y)}{C'^2(y)} = -\frac{\theta(x,y)^2 C''(y)}{\delta(x)}. \tag{C.4}$$

By Assumption 7, $\lim_{y\downarrow\inf\mathcal{Y}_{x,s}}\theta(x,y)^2 f_{\theta|X=x,S=s}(\theta(x,y)) = 0$. Because C'' is bounded, this implies that $\lim_{y\downarrow\inf\mathcal{Y}_{x,s}}f_{Y|X=x,S=s}(y) = 0$. Hence, $f_{Y|X=x,S=s}$ is continuous or can be extended by continuity on $[0, \sup\mathcal{Y}_{x,s}]$.

Similarly, $\lim_{y\uparrow\sup\mathcal{Y}_{x,s}}C''(y)/C'^2$ exists. Hence, $\lim_{y\uparrow\sup\mathcal{Y}_{x,s}}f_{Y|X=x,S=s}(y)$ exists as well. If $\underline{\theta} > 0$, $f_{\theta|X=x,S=s}(\underline{\theta}) = 0$ and so $\lim_{y\uparrow\sup\mathcal{Y}_{x,s}}f_{Y|X=x,S=s}(y) = 0$. If $\underline{\theta} = 0$, $\sup\mathcal{Y}_{x,s} = 1$. Hence, in both cases we can extend by continuity $f_{Y|X=x,S=s}$ on $[0, 1]$.

Let $\mathcal{F}$ denote the space of continuous density functions on $[0, 1]$. For $f \in \mathcal{F}$, $n \in \mathbb{N}$ and $r \in \{0, \dots, n\}$, let

$$\ell(f, r, n) = \ln\left(\int_0^1 y^r(1-y)^{n-r}f(y)\,dy\right),$$

let $Q_{x,s}(f) = E(\ell(f, R, N)|X = x, S = s)$, and

$$\widehat{Q}_{x,s}(f) = \sum_{i:X_i=x,S_i=s}\ell(f, R_i, N_i).$$

By definition of $\widehat{f}_{Y|X=x,S=s}$, $\widehat{f}_{Y|X=x,S=s} = \arg\max_{f\in\mathcal{F}_L}\widehat{Q}_{x,s}(f)$ is a sieve M-estimator. We use Theorem 3.1 of Chen (2007) and its associated Remark 3.2 to prove (C.3). To this end, we check the following conditions:

a. $Q_{x,s}$ is uniquely maximized at $f_{Y|X=x,S=s}$ and $Q_{x,s}(f_{Y|X=x,S=s}) > -\infty$.

b. For all L, $\mathcal{F}_L \subset \mathcal{F}_{L+1}$ and for all $f \in \mathcal{F}$, there exists $f_L \in \mathcal{F}_L$ such that $\|f_L - f\| \to 0$.

c. $Q_{x,s}$ is continuous for $\|\cdot\|$.

d. $\mathcal{F}_L$ is compact.

e. $E[\sup_{f\in\mathcal{F}_L}|\ell(f, R, N)||X = x, S = s] < \infty$.

f. There exists $U(\cdot, \cdot)$ such that $E(U(R, N)|X = x, S = s) < \infty$ and for all $(f, g) \in \mathcal{F}_L^2$, $|\ell(f, R, N) - \ell(g, R, N)| \le \|f - g\|U(R, N)$.

g. The minimal number of δ-balls that cover $\mathcal{F}_L$, denoted $N_b(\delta, \mathcal{F}_L, \|\cdot\|)$, satisfies $\ln N_b(\delta, \mathcal{F}_L, \|\cdot\|) = o(L)$.

a. First, for all $g \in \mathcal{F}$,

$$E\left[\frac{\exp\ell(g, R, N)}{\exp\ell(f_{Y|X=x,S=s}, R, N)}\Big|N = n, X = x, S = s\right]$$

$$= \sum_{r=0}^n \Pr(R = r|N = n, X = x, S = s)\frac{\binom{r}{n}\int_0^1 y^r(1-y)^{n-r}g(y)\,dy}{\Pr(R = r|N = n, X = x, S = s)}$$

$$= \int_0^1 \left(\sum_{r=0}^n \binom{r}{n} y^r (1-y)^{n-r} \right) g(y)\, dy$$

$$= \int_0^1 g(y)\, dy$$

$$= 1.$$

Thus,

$$E\left[\frac{\exp \ell(g, R, N)}{\exp \ell(f_{Y|X=x,S=s}, R, N)} \Big| X = x, S = s \right] = 1.$$

Besides, because $f_{Y|X=x,S=s}$ is identified, we have $\ell(g, R, N) \neq \ell(f_{Y|X=x,S=s}, R, N)$ with a strictly positive probability for all $g \neq f_{jpj}$. Thus, by Jensen's inequality,

$$E\left[\ln\left(\frac{\exp \ell(g, R, N)}{\exp \ell(f_{Y|X=x,S=s}, R, N)} \right) \Big| X = x, S = s \right]$$

$$< \ln E\left[\frac{\exp \ell(g, R, N)}{\exp \ell(f_{Y|X=x,S=s}, R, N)} \Big| X = x, S = s \right] = 0.$$

This proves that $Q_{x,s}$ is uniquely maximized at $f_{Y|X=x,S=s}$. Moreover, let $u_1 \in (0, 1)$ be such that $\int_{u_1}^{1-u_1} f_{Y|X=x,S=s}(y)\, dy \geq 1/2$. We have

$$\int_0^1 y^r (1-y)^{n-r} f_{Y|X=x,S=s}(y)\, dy \geq \int_{u_1}^{1-u_1} y^r (1-y)^{n-r} f_{Y|X=x,S=s}(y)\, dy$$

$$\geq u_1^n \int_{u_1}^{1-u_1} \left(\frac{y}{u_1} \right)^r \left(\frac{1-y}{u_1} \right)^{n-r} f_{Y|X=x,S=s}(y)\, dy$$

$$\geq u_1^n \int_{u_1}^{1-u_1} f_{Y|X=x,S=s}(y)\, dy$$

$$\geq \frac{u_1^n}{2}. \tag{C.5}$$

As a result, $Q_{x,s}(f_{Y|X=x,S=s}) \geq E(N|X = x, S = s) \ln u_1 - \ln 2$. By Assumption 7, $E(N|X = x, S = s) < \infty$, so that $Q_{x,s}(f_{Y|X=x,S=s}) > -\infty$.

b.   First, $\mathcal{F}_L \subset \mathcal{F}_{L+1}$ for all N since K_L is increasing. Now fix $f \in \mathcal{F}$ and $\varepsilon > 0$. Because $\sqrt{f}$ is continuous on $[0, 1]$, there exists, by Weierstrass' theorem, a polynomial P of order J such that $\|\sqrt{f} - P\| \leq \varepsilon$. Then

$$\|f - P^2\| \leq \|\sqrt{f} - P\| \times \|\sqrt{f} + P\|$$

$$\leq \|\sqrt{f} - P\| \times \left(2\|\sqrt{f}\| + \|P - \sqrt{f}\| \right)$$

$$\leq \varepsilon \left(\varepsilon + 2\|\sqrt{f}\| \right).$$

Now let N be such that $K_L \geq 2J$ and

$$M \ln K_L \geq \frac{\varepsilon(\varepsilon + 2\|\sqrt{f}\|) + \|\sqrt{f}\|}{1 - \varepsilon(\varepsilon + 2\|\sqrt{f}\|)}.$$

We have

$$\int_0^1 P^2(y)\, dy \geq \int_0^1 f(y)\, dy - \int_0^1 |f(y) - P^2(y)|\, dy \geq 1 - \varepsilon(\varepsilon + 2\|\sqrt{f}\|).$$

Thus, defining $f_L = P^2/(\int_0^1 P^2(y)\, dy)$, we get

$$\begin{aligned}
\|f_L\| &\leq \frac{\|P^2\|}{1 - \varepsilon(\varepsilon + 2\|\sqrt{f}\|)} \\
&\leq \frac{\|P^2 - f\| + \|f\|}{1 - \varepsilon(\varepsilon + 2\|\sqrt{f}\|)} \\
&\leq \frac{\varepsilon(\varepsilon + 2\|\sqrt{f}\|) + \|\sqrt{f}\|}{1 - \varepsilon(\varepsilon + 2\|\sqrt{f}\|)} \\
&\leq M \ln K_L,
\end{aligned}$$

so that $f_L \in \mathcal{F}_L$. Moreover,

$$\begin{aligned}
\|f - f_L\| &\leq \|f - P^2\| + \|P^2\| \left| 1 - \frac{1}{\displaystyle\int_0^1 P^2(u)\, du} \right| \\
&\leq \varepsilon(\varepsilon + 2\|\sqrt{f}\|) + (\|f\| + \varepsilon(\varepsilon + 2\|\sqrt{f}\|)) \left(\frac{1}{1 - \varepsilon(\varepsilon + 2\|\sqrt{f}\|)} - 1 \right).
\end{aligned}$$

This establishes b, since the right-hand side tends to zero with ε.

c. Fix $\varepsilon > 0$ and $f \in \mathcal{F}$ and let $g \in \mathcal{F}$ be such that $\|f - g\| \leq \varepsilon$. For all $n \in \mathbb{N}$ and $r \in \{0, \dots, n\}$,

$$\left| \int_0^1 y^r (1-y)^{n-r} f(y)\, dy - \int_0^1 y^r (1-y)^{n-r} g(y)\, dy \right| \leq \|f - g\| \leq \varepsilon. \tag{C.6}$$

Moreover, there exists $u_2 \in (0, 1)$ such that

$$\int_{u_2}^{1-u_2} f(y)\, dy \wedge \int_{u_2}^{1-u_2} g(y)\, dy \geq \frac{1}{2}.$$

Hence, reasoning as in (C.5), we get

$$\int_0^1 y^r (1-y)^{n-r} f(y)\, dy \wedge \int_0^1 y^r (1-y)^{n-r} g(y)\, dy \geq \frac{u_2^n}{2}. \tag{C.7}$$

Besides, for all $a, b > 0$, $|\ln b - \ln a| \le |b - a|/a \wedge b$. Hence, using (C.6) and (C.7), we get, for all $n \in \mathbb{N}$ and $r \in \{0, \ldots, n\}$,

$$
\begin{aligned}
\left| \ell(f, r, n) - \ell(g, r, n) \right| &= \left| \ln\left(\int_0^1 y^r (1-y)^{n-r} f(y) \, dy \right) - \ln\left(\int_0^1 y^r (1-y)^{n-r} g(y) \, dy \right) \right| \\
&\le \frac{\left| \int_0^1 y^r (1-y)^{n-r} f(y) \, dy - \int_0^1 y^r (1-y)^{n-r} g(y) \, dy \right|}{\left(\int_0^1 y^r (1-y)^{n-r} f(y) \, dy \right) \wedge \left(\int_0^1 y^r (1-y)^{n-r} g(y) \, dy \right)} \\
&\le \frac{2\varepsilon}{u_2^n}. \tag{C.8}
\end{aligned}
$$

As a result,

$$
\left| Q_{x,s}(f) - Q_{x,s}(g) \right| \le E\left[\left| \ell(f, R, N) - \ell(g, R, N) \right| \right] \le 2\varepsilon E\left(\frac{1}{u_2^n} \Big| X = x, S = s \right).
$$

The expectation is finite by Assumption 7. Hence, $Q_{x,s}$ is continuous for $\| \cdot \|$.

d. $\mathcal{F}_L$ is closed, bounded, and belongs to a finite dimensional space. $\mathcal{F}_L$ is thus compact.

e. Because $|g(x)| \le M \ln K_L$ for all $g \in \mathcal{F}_L$, there exists $u_3 \in (0, 1/2)$ such that for all $g \in \mathcal{F}_L$, $\int_{u_3}^{1-u_3} g(y) \, dy \ge 1/2$. Reasoning as previously, we have

$$
m(n, r) = \inf_{g \in \mathcal{F}_L} \int_0^1 y^r (1-y)^{n-r} g(y) \, dy \ge \frac{u_3^n}{2}. \tag{C.9}
$$

Besides, for all $f \in \mathcal{F}_L$, $n \in \mathbb{N}$ and $r \in \{0, \ldots, n\}$,

$$
\begin{aligned}
\left| \ell(f, r, n) \right| &= \left| \ln \int_0^1 y^r (1-y)^{n-r} f(y) \, dy \right| \\
&\le \left| \ln\left(\inf_{g \in \mathcal{F}_L} \int_0^1 y^r (1-y)^{n-r} g(y) \, dy \right) \right|.
\end{aligned}
$$

Thus,

$$
E\left[\sup_{f \in \mathcal{F}_L} \left| \ell(f, R, N) \right| \Big| X = x, S = s \right] \le E\left[\left| \ln m(N, R) \right| \Big| X = x, S = s \right] \tag{C.10}
$$

$$
\le E\left[\left| \ln 2 \right| + N \left| \ln u_3 \right| \Big| X = x, S = s \right], \tag{C.11}
$$

and $E(N | X = x, S = s) < \infty$ implies that $E[\sup_{f \in \mathcal{F}_L} |\ell(f, N)| | X = x, S = s] < \infty$.

f. Using (C.9) and a similar argument as in (C.8), we get, for all $(f, g) \in \mathcal{F}_L$,

$$
\left| \ell(f, R, N) - \ell(g, R, N) \right| \le \frac{2 \| f - g \|}{u_3^N}.
$$

Thus, by Assumption 7, Point f is satisfied with $U(r, n) = 2/u_3^n$.

g. For all $f \in \mathcal{F}_L$ by Markov's inequality on polynomials (see, e.g., Borwein and Erdélyi (1995, Theorem 5.1.8)),

$$\|f'\| \leq 2(2K_L)^2 \|f\| \leq 8MK_L^2 \ln K_L.$$

$\mathcal{F}_L$ is thus included in the set

$$\mathcal{G}_L = \{f : \forall (x, y) \in [0, 1]^2, |f(x)| \leq M \ln K_L, |f(x) - f(y)| \leq 8MK_L^2 \ln K_L\}.$$

This set is a particular case of a more general class considered by van der Vaart and Wellner (1996, Theorem 2.7.1). They prove that there exists a constant $C_0 > 0$ such that

$$\ln N_b(\delta, \mathcal{G}_L, \|\cdot\|) \leq C_0 K_L^2 \ln K_L.$$

Because $\ln N_b(\delta, \mathcal{F}_L, \|\cdot\|) \leq \ln N_b(\delta, \mathcal{G}_L, \|\cdot\|)$ and $K_L^2 \ln K_L / L \to 0$, $\ln N_b(\delta, \mathcal{F}_L, \|\cdot\|) = o(L)$, which ends the proof of (C.3).

2. Uniform consistency of $\widehat{H}$.

We now establish that for all $(\underline{x}, \overline{x})$ such that $I \equiv [\underline{x}, \overline{x}] \subset \overset{\circ}{\mathcal{Y}}_1$,

$$\sup_{x \in I} |\widehat{H}(x) - H(x)| \overset{\mathbb{P}}{\longrightarrow} 0. \tag{C.12}$$

By Assumption 2, $\partial\theta/\partial(2, y)y < 0$ and $f_\theta(\theta(2, y)) > 0$ for all $y \in (0, 1)$. Hence, by continuity of $f_{\theta|S=1}$ and $\partial\theta(2, \cdot)/\partial y$, for all compact K strictly included in $(0, 1)$,

$$\min_{y \in K} f_{Y|X=2,S=1}(y) = \min_{y \in K} \left[-f_{\theta|S=1}(\theta(2, y)) \partial\theta(2, y)/\partial y \right] > 0. \tag{C.13}$$

Now, consider the mapping $(F, G) \mapsto F^{-1} \circ G$, for any strictly increasing cdf F and where G is any continuous function defined on I such that $G(I) \subset (0, 1)$. By the chain rule and Hadamard differentiability of the quantile function (see, respectively, Theorem 20.9 and Lemma 21.4 of van der Vaart (1998)), it is Hadamard differentiable and, therefore, continuous, at $(F_{Y|X=2,S=1}, F_{Y|X=1,S=1})$. By Step 1 of the proof, $(\widehat{F}_{Y|X=2,S=1}, \widehat{F}_{Y|X=1,S=1})$ converges in probability uniformly to $(F_{Y|X=2,S=1}, F_{Y|X=2,S=1})$. Hence, by the continuous mapping theorem (see, e.g., Theorem 18.11 in van der Vaart (1998)), (C.12) holds.

3. Consistency of $\widehat{\mathcal{K}}$ and $\widehat{y}_k$, for all $k \in \mathcal{K}$.

We now prove that for all $k \in \mathcal{K}$ and for all $\varepsilon > 0$, as $N \to \infty$,

$$\Pr(k \in \widehat{\mathcal{K}}, |\widehat{y}_k - y_k| \leq \varepsilon) \to 1. \tag{C.14}$$

Let us proceed by induction on $k \in \mathcal{K}$. The proposition is true when $k = 0$. Suppose that it holds for $k - 1 \geq 0$ and let us prove that it holds for k if $k \in \mathcal{K}$ (the proof is similar for negative values). First,

$$\begin{aligned}
\widehat{F}_{Y|X=1,S=1}&(\widehat{y}_{k-1}) \\
&\leq \|\widehat{F}_{Y|X=1,S=1} - F_{Y|X=1,S=1}\| + |F_{Y|X=1,S=1}(\widehat{y}_{k-1}) - F_{Y|X=1,S=1}(y_{k-1})| \\
&\quad + F_{Y|X=1,S=1}(y_{k-1}).
\end{aligned}$$

The first term on the right-hand side converges to 0 with probability approaching one (w.p.a.o.) by Step 1. The second term also converges to 0 in probability by the induction hypothesis and continuity of $F_{Y|X=1,S=1}$. Finally, by assumption, $y_{k-1} \in \overset{\circ}{\mathcal{Y}}_1$, implying that $F_{Y|X=1,S=1}(y_{k-1}) < 1 - \tau_L$ for L large enough. Hence, w.p.a.o., $\widehat{F}_{Y|X=1,S=1}(\widehat{y}_{k-1}) \leq 1 - \tau_L$, implying that $k \in \widehat{\mathcal{K}}$.

Next, by the induction hypothesis and $y_{k-1} \in \overset{\circ}{\mathcal{Y}}_1$, there exists an interval $I \subset \overset{\circ}{\mathcal{Y}}_1$ such that w.p.a.o., $\widehat{y}_{k-1} \in I$ and $k \in \widehat{\mathcal{K}}$. Under this event,

$$|\widehat{y}_k - y_k| \leq \sup_{x \in I} |\widehat{H}(x) - H(x)| + |H(\widehat{y}_{k-1}) - H(y_{k-1})|.$$

Therefore, by Step 2 and continuity of H, (C.14) holds, which concludes the induction step. Hence, (C.14) holds for all $k \in \mathcal{K}$.

We have proved so far that if $k \in \mathcal{K}$, $k \in \widehat{\mathcal{K}}$ w.p.a.o. Now let us prove the reverse, namely that if $k \notin \mathcal{K}$, $k \notin \widehat{\mathcal{K}}$ w.p.a.o. We focus again on $k > 0$, the proof being similar for negative values. Given the definition of $\mathcal{K}$ and $\widehat{\mathcal{K}}$, it suffices to consider the case where $\overline{k} \equiv \sup \mathcal{K} < +\infty$, and show that $\overline{k} + 1 \notin \widehat{\mathcal{K}}$ w.p.a.o. First, $y_{\overline{k}} \notin \mathcal{Y}_1$. By assumption, $y_k > \sup \mathcal{Y}_1$. Hence, in view of what precedes, we also have $\widehat{y}_{\overline{k}} > \sup \mathcal{Y}_1$ and thus $F_{Y|X=1,S=1}(\widehat{y}_{\overline{k}}) = 1$ w.p.a.o. Then w.p.a.o.,

$$\widehat{F}_{Y|X=1,S=1}(\widehat{y}_{\overline{k}}) = \widehat{F}_{Y|X=1,S=1}(\widehat{y}_{\overline{k}}) - F_{Y|X=1,S=1}(\widehat{y}_{\overline{k}}) + F_{Y|X=1,S=1}(\widehat{y}_{\overline{k}})$$

$$\geq 1 - \|\widehat{F}_{Y|X=1,S=1} - F_{Y|X=1,S=1}\|$$

$$> 1 - \tau_L,$$

the latter in view of the condition on τ_L. Therefore, $\widehat{y}_k \notin \widehat{Y}_1$, and $\overline{k} + 1 \notin \widehat{\mathcal{K}}$.

4. Consistency of the estimators of the bounds of C' and $F_{\theta|X,S}$.

Let us first prove that $\widehat{\underline{C}}'(y)$ is consistent for every $y \notin \{y_k, k \in \mathcal{K}\}$. First, suppose that $\underline{C}'(y) = 0$. Then $y_k > y$ for all $k \in \mathcal{K}$. Then $\underline{k} = \inf \mathcal{K} > -\infty$; otherwise the sequence $(y_k)_{-k \in \mathbb{N}}$ would tend to zero, implying that $\underline{C}'(y) \geq C'(y_k) > 0$ for some k. Also, $\underline{k} \in \mathcal{K}$ and $\underline{k} - 1 \notin \mathcal{K}$. By Step 3, $\underline{k} \in \widehat{\mathcal{K}}$ and $\underline{k} - 1 \notin \widehat{\mathcal{K}}$ w.p.a.o. Hence, $\underline{k} = \inf \widehat{\mathcal{K}}$ w.p.a.o. Because $y_{\underline{k}} > y$ and since $\widehat{y}_{\underline{k}}$ is consistent by Step 3, $\widehat{y}_k > y$ for all $k \in \widehat{\mathcal{K}}$, which implies in turn that $\widehat{\underline{C}}'(y) = 0$ w.p.a.o. Now, suppose that $\underline{C}'(y) > 0$. Then, by definition of $\underline{C}'(y)$ and because $y \notin \{y_k, k \in \mathcal{K}\}$, there exists $\overline{k}(y) \in \mathcal{K}$ such that $y_{\overline{k}(y)} < y$ and either $\overline{k}(y) + 1 \in \mathcal{K}$, in which case $y < y_{\overline{k}(y)+1}$, or $\overline{k}(y) + 1 \notin \mathcal{K}$. In the former case, $(\overline{k}(y), \overline{k}(y) + 1) \in \widehat{\mathcal{K}}^2$ and $\widehat{y}_{\overline{k}(y)} < y < \widehat{y}_{\overline{k}(y)+1}$ w.p.a.o. In the latter case, $\widehat{y}_{\overline{k}(y)} < y$ and $\overline{k}(y) + 1 \notin \widehat{\mathcal{K}}$ w.p.a.o. In both cases, $\widehat{\underline{C}}'(y) = \underline{C}'(y)$.

The consistency of $\widehat{\overline{C}}'(y)$ for $y \notin \{y_k, k \in \mathcal{K}\}$ follows similarly. Now, let us turn to $\widehat{\underline{F}}_{\theta|X=x,S=s}(\theta)$. If $\underline{F}_{\theta|X=x,S=s}(\theta) = 0$, then $\theta_k > \theta$ for all k such that $k + \mathbb{1}\{x = 2\} \in \mathcal{K}$. Because $(\theta_k)_{k \in \mathbb{N}}$ tends to 0, $\overline{k} \equiv \sup \mathcal{K} < +\infty$. Also, $\overline{k} \in \mathcal{K}$ and $\overline{k} + 1 \notin \mathcal{K}$. Then w.p.a.o., $\overline{k} \in \widehat{\mathcal{K}}$ and $\overline{k} + 1 \notin \widehat{\mathcal{K}}$, implying that $\widehat{\underline{F}}_{\theta|X=x,S=s}(\theta) = 0$ w.p.a.o. If $\underline{F}_{\theta|X=x,S=s}(\theta) > 0$, there exists $\overline{k}(\theta)$ such that $k + \mathbb{1}\{k = 2\} \in \mathcal{K}$, $\theta_{\overline{k}(\theta)} \leq \theta$ and either $\theta < \theta_{\overline{k}(\theta)+1}$ or $\overline{k}(\theta) + 1 + \mathbb{1}\{k = 2\} \notin \mathcal{K}$.

W.p.a.o., the same holds with $\mathcal{K}$ replaced by $\widehat{\mathcal{K}}$. Then, w.p.a.o.,

$$\widehat{\underline{F}}_{\theta|X=x,S=s}(\theta) = 1 - \widehat{F}_{Y|X=x,S=s}(\widehat{\underline{y}}_{\overline{k}(\theta)}).$$

Hence, by uniform convergence of $\widehat{F}_{Y|X=x,S=s}$ and convergence of $\widehat{\underline{y}}_{\overline{k}(\theta)}$ and continuity of $F_{Y|X=x,S=s}$, $\widehat{\underline{F}}_{\theta|X=x,S=s}(\theta)$ converges to $\underline{F}_{\theta|X=x,S=s}(\theta)$. The same reasoning applies to the upper bound. Finally, for any $k \in \mathcal{K}$, $k \in \widehat{\mathcal{K}}$ w.p.a.o. and by definition,

$$\underline{\widehat{C}}'(\widehat{y}_k) = \overline{\widehat{C}}'(\widehat{y}_k) = C'(y_k).$$

This proves the last assertion of the theorem.   □

PROOF OF PROPOSITION 4.1. If C' is identified, $\lambda(x)$ can be recovered with (4.11). The current surplus can then be obtained directly using equation (4.4). Similarly, under Assumption 2, $y_n(x, \theta, t)$ defined by (4.5) only depends on C' and on the distribution of θ conditional on P. It is thus identified if the latter are identified, implying that the optimal contract given by (4.6), and the corresponding surplus, is identified. In the model under complete information, the probabilities $y^C(x, \theta)$ fixed under complete information are identified from (4.9), as long as C' is fully identified. Because we have supposed that Insee fixes $w(x)$ so as to ensure universal participation, we can recover the optimal transfer function $t_n^C(x, \cdot)$ and the corresponding surplus from equation (4.10) and $t_n^C(x, 0) = nw(x)$.   □

REFERENCES

Abbring, J. H. and G. Ridder (2015), "Regular variation and the identification of generalized accelerated failure-time models." *Econometric Theory*, 31, 1229–1248. [351, 362]

Aryal, G. and M. Gabrielli (2018), "An empirical analysis of competitive nonlinear pricing." SSRN Working Paper, https://ssrn.com/abstract=2262664. [351]

Aryal, G., I. Perrigne, and Q. Vuong (2016), "Identification of insurance models with multidimensional screening." ArXiv e-prints, https://arxiv.org/abs/1508.02919. [351]

Borwein, P. B. and T. Erdélyi (1995), *Polynomials and Polynomials Inequalities*. Springer-Verlag. [392]

Campanelli, P. C., E. A. Martin, and J. M. Rothgeb (1991), "The use of respondent and interviewer debriefing studies as a way to study response error in survey data." *Journal of the Royal Statistical Society: Series D (The Statistician)*, 40, 253–264. [381]

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models." In *Handbook of Econometrics*, Vol. 6 (J. J. Heckman and E. E. Leamer, eds.), North Holland. [364, 388]

Chiappori, P. A. and B. Salanié (2003), "Testing contract theory: A survey of some recent work." In *Advances in Economics and Econometrics*, Vol. 1 (L. H. e. S. T. M. Dewatripont, ed.), Cambridge University Press, New York, NY. [352]

Chu, C. S., P. Leslie, and A. Sorensen (2011), "Bundle-size pricing as an approximation to mixed bundling." *The American Economic Review*, 101, 263–303. [378]

Chu, L. Y. and D. E. M. Sappington (2007), "Simple cost-sharing contracts." *The American Economic Review*, 97, 419–428. [351, 378]

Chung, D., T. J. Steenburgh, and K. Sudhir (2014), "Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans." *Marketing Science*, 33, 165–187. [350]

Cleary, P. D., D. Mechanic, and N. Weiss (1981), "The effect of interviewer characteristics on responses to a mental health interview." *Journal of Health and Social Behavior*, 22, 183–193. [381]

Copeland, A. and C. Monnet (2009), "The welfare effect of incentive schemes." *Review of Economic Studies*, 76, 93–113. [350]

Crawford, G. S. and M. Shum (2007), "Monopoly quality degradation and regulation in cable television." *Journal of Law and Economics*, 50, 181–219. [351]

D'Haultfœuille, X. and P. Février (2010), "Identification of a class of adverse selection models with contracts variation." CREST Working Paper, https://ideas.repec.org/p/crs/wpaper/2011-27.html. [357, 363]

D'Haultfœuille, X. and R. Rathelot (2017), "Measuring segregation on small units: A partial identification analysis." *Quantitative Economics*, 8 (1), 39–73. [358, 383, 384]

Ferrall, C. and B. Shearer (1999), "Incentives and transaction costs within the firm: Estimating an agency model using payroll records." *Review of Economic Studies*, 99, 309–338. [350, 351, 352, 357, 377, 381]

Gagnepain, P. and M. Ivaldi (2002), "Incentive regulatory policies: The case of public transit systems in France." *The RAND Journal of Economics*, 33, 605–629. [350]

Gagnepain, P., M. Ivaldi, and D. Martimort (2013), "The cost of contract renegotiation: Evidence from the local public sector." *American Economic Review*, 103, 2352–2383. [351]

Guerre, E., I. Perrigne, and Q. Vuong (2009), "Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions." *Econometrica*, 77, 1193–1227. [351, 363]

Gut, A. (2005), *Probability: A Graduate Course*. Springer-Verlag, New York, NY. [358, 385]

Holmstrom, L. and P. Milgrom (1987), "Aggregation and linearity in the provision of intertemporal incentives." *Econometrica*, 55, 303–328. [351]

Ivaldi, M. and D. Martimort (1994), "Competition under nonlinear pricing." *Annales d'Economie et de Statistique*, 34, 71–114. [350]

Kang, K. and B. Silveira (2018), "Understanding disparities in punishment: Regulator preferences and expertise." Working Paper. [351]

Laffont, J. J. and D. Martimort (2002), *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press. [351]

Laffont, J. J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*. MIT Press. [357]

Lavergne, P. and A. Thomas (2005), "Semiparametric estimation and testing in a model of environmental regulation with adverse selection." *Empirical Economics*, 30, 171–192. [350, 351, 357]

Lazear, E. (2000), "Performance pay and productivity." *American Economic Review*, 90, 1346–1361. [349, 372]

Le Lan, R. (2009), "Enquêtes ménages de l'insee: vers la fin de la baisse des taux de réponse?" *Courrier des Statistiques*, 128, 33–41. [359]

Leslie, P. (2004), "Price discrimination in broadway theater." *The RAND Journal of Economics*, 35, 520–541. [350]

Lim, C. and A. Yurukoglu (2018), "Structural analysis of nonlinear pricing." *Journal of Political Economy*, 126, 2523–2568. [351]

Luo, Y., I. Perrigne, and Q. Vuong (2018), "Structural analysis of nonlinear pricing." *Journal of Political Economy*, 126, 2523–2568. [351]

Mensh, B. S. and D. B. Kandel (1988), "Underreporting of substance use in a national longitudinal youth cohort individual and interviewer effects." *The Public Opinion Quarterly*, 52, 100–124. [381]

Miravete, E. J. (2002), "Estimating demand for local telephone service with asymmetric information and optional calling plans." *Review of Economic Studies*, 69, 943–971. [350]

Miravete, E. J. (2007), "The limited gains from complex tariffs." Working Paper Series 3971, Victoria University of Wellington, The New Zealand Institute for the Study of Competition and Regulation, https://ideas.repec.org/p/vuw/vuwcsr/3971.html. [351, 378]

Miravete, E. J. and L.-H. Roller (2005), "Estimating markups under nonlinear pricing competition." *Journal of the European Economic Association*, 2, 526–535. [350]

Neeman, Z. (2003), "The effectiveness of English auctions." *Games and Economic Behavior*, 43, 214–238. [378]

O'Muircheartaigh, C. and P. Campanelli (1998), "The relative impact of interviewer effects and sample design effects on survey precision." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161, 63–77. [381]

Paarsch, H. and B. Shearer (2000), "Piece rates, fixed wages and incentive effects: Statistical evidence from payroll records." *International Economic Review*, 41, 59–92. [350]

Perrigne, I. and Q. Vuong (1999), "Structural econometrics of first-price auctions: A survey of methods." *Canadian Journal of Agricultural Economics*, 47, 203–223. [352]

Perrigne, I. and Q. Vuong (2011a), "Nonlinear pricing in yellow pages." Report. [351]

Perrigne, I. and Q. Vuong (2011b), "Nonparametric identification of a contract model with adverse selection and moral hazard." *Econometrica*, 79, 1499–1539. [351]

Prendergast, C. (1999), "The provision of incentives in firms." *Journal of Economic Literature*, 37, 7–63. [349, 372, 373]

Raju, J. S. and V. Srinivasan (1996), "Quota-based compensation plans for multiterritory heterogeneous salesforces." *Management Science*, 6, 1454–1462. [351]

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. [351, 378]

Singer, E., M. R. Frankel, and M. B. Glassman (1983), "The effect of interviewer characteristics and expectations on response." *Public Opinion Quarterly*, 47, 68–83. [381]

van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. [392]

van der Vaart, A. W. and J. Wellner (1996), *Weak Converence and Empirical Process*. Springer. [392]

Wilson, R. (1993), *Nonlinear Pricing*. Oxford University Press. [351, 357, 378]

Wolak, F. (1994), "An econometric analysis of the asymmetric information, regulator-utility interaction." *Annales d'Economie et de Statistique*, 34, 13–69. [350, 357]

Wood, G. R. (1999), "Binomial mixtures: Geometric estimation of the mixing distribution." *Annals of Statistics*, 27, 1706–1721. [358]