# The Provision of Wage Incentives: A Structural Estimation Using Contracts Variation*

Xavier D'Haultfœuille [†] Philippe Février[‡]

July 2015

**Abstract**

We address the issues of the optimality of simple linear compensation contracts and the importance of asymmetries between firms and workers in the context of contracts between the French National Institute of Statistics and Economics (Insee) and the interviewers it hires 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 numbers: C14, D82, D86**

*We would like to thank Gaurab Aryal, Steve Berry, Raicho Bojilov, Thierry Magnac, Arnaud Maurel, Aviv Nevo, Martin Pesendorfer, Rob Porter, Jean-Marc Robin, Jean-Charles Rochet, Mathieu Rosenbaum, Bernard Salanié, Michael Visser and the participants of various seminars and conferences for helpful discussions and comments. We finally acknowledge Daniel Verger and the Unité Méthodes Statistiques of Insee for providing us with the data.

[†]CREST. E-mail address: xavier.dhaultfoeuille@ensae.fr.

[‡]CREST. E-mail address: fevrier@ensae.fr.

## 1 Introduction

Over the past three decades, extensive attention has been devoted to asymmetries of information and their consequences in economics. 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 & Shearer, 2000), and that the form of the payment scheme matters (Ferrall & Shearer, 1999, Chung et al., 2013 or Copeland & Monnet, 2009). 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 which 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) which 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 their heterogeneity, which is the reason why Insee faces an asymmetric information problem. This asymmetry of information is due to differences between interviewers themselves and to differences between the geographical areas in which they are working. In response to this asymmetric information problem, and to give incentives to its interviewers, Insee 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 nonparametrically identified using the exogenous change in contracts.[1] 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, nonlinear and price discrimination models.[2] All these models share a common underlying structure for which our procedure 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 & Vuong (2011), Aryal et al. (2010), Luo et al. (2014) and Aryal (2014) on respectively regulation, insurance models, unidimensional and multidimensional nonlinear pricing, respectively. It is also related to the identification of first-price auctions models with risk-adverse bidders, using exogenous variations in the number of bidders (Guerre et al., 2009). Finally, it is related to the identification of nonseparable models with discrete instruments (see D'Haultfœuille & Février, 2015 and Torgovitsky, 2015).

Beyond identification, we also develop a nonparametric estimation procedure using our identification method. We thus recover nonparametrically some points 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 & Martimort, 2002, for a survey),[3] simple compensation schemes such as piece rates and bonuses are usually thought of as the best compromise between efficiency and ease of implementation

[1]In a related paper, we study the identification of adverse selection models under more general exogenous changes (see D'Haultfœuille & Février, 2010).

[2]For an incomplete list of empirical papers in these fields, see Ivaldi & Martimort (1994),Wolak (1994), Gagnepain & Ivaldi (2002), Miravete (2002), Perrigne (2002), Leslie (2004), Miravete & Roller (2005), Lavergne & Thomas (2005), Crawford & Shum (2007), Perrigne & Vuong (2010), Miravete (2007) and Gagnepain et al. (2013).

[3]An exception is the result of Holmstrom & Milgrom (1987).

(Raju & 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 & 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 & Shearer (1999), on the other hand, concludes that simple nonlinear compensation plans lead to substantial inefficiencies.

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 & Salanié, 2002, for a survey), or in the auction literature (see Perrigne & 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 & 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. All proofs are deferred to the appendix.

## 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. To conduct the interviews, Insee uses interviewers, whose 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 mandatory to answer Insee questionnaires. In practice, more than 90% of households indeed participate. In a typical household survey, it takes around one hour to go through all the questions. In compensation, interviewers are paid in a similar way for all household surveys. They receive a basic wage for each household they have to interview, plus a bonus for each interview they achieve. They are 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 have its characteristics in the 1999 census (namely, the number of rooms, the household size and the age of the reference person), 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% respectively). 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

[4] We also have some limited information on interviewers that we use at the end of our analysis, see Appendix B 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

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 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]

| 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% |

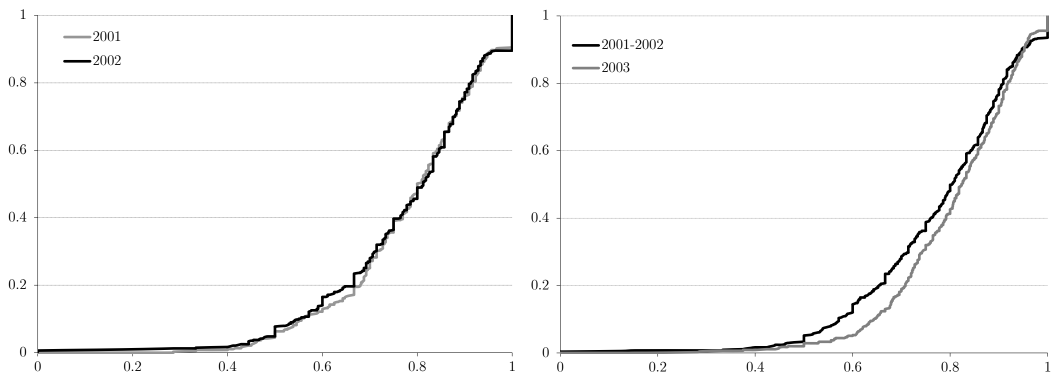



Figure 1: Distribution functions of the response rates on all interviewers, for all households.

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 larger in 2003 than 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 subsample of families were drawn similarly.

the total number of households, where the 2001 and 2002 data are pooled.

[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.

Table 2 shows that, as expected, the average response rates for families are higher than in the population in general (respectively 79.0%, 79.8% and 83.1% versus 78.5%, 77.7% and 80.7%). Comparing the statistics of the three surveys, we find, however, the same patterns as the one found in Table 1. There is no significant difference between the 2001 and 2002 surveys (79.0% and 79.8% respectively) whereas interviewers achieve significantly higher response rates in 2003 (83.1%).

| | Number of | Number of | Average | Payment per household | | Average income | | |
|---|---|---|---|---|---|---|---|---|
| Year | interviewers | families | response rate | 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 |

Table 2: Descriptive statistics on the subsample of families.

There is also a second difference in the three surveys, namely their payment schemes. Whereas the basic wage is nearly constant the three 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

To analyze the issue of contracts optimality and quantify the loss due to asymmetric information and linear contracts, we first model the interviewers' decision, in particular to recover their utility function. We will use this utility function in the next section to perform counterfactual analysis.

### 3.1 The interviewers' program

We suppose that the interviewer decides what effort to 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 heterogenous and may be easy or difficult to contact, depending on their characteristics. It is, for instance, difficult to contact a single person in a urban area. Indeed, single persons living in urban areas spend relatively little time at home, and digital locks for instance make a direct contact more difficult to establish.

[7] All figures are in 2002 euros.

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 homogenous 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 homogenous 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, $y$ is not equal to the actual response rate because of randomness in interviewers' work. The expectation of the cost for interviewer $i$ to reach a probability $y$ in survey $j \in \{0, 1, 2\}$ (corresponding to the 2001, 2002 and 2003 surveys, respectively), with $n_{ij} = n$ households to interview, is denoted by $C_{ijn}(y)$. Note that an interviewer may not participate to a survey $j$. We consider that this case corresponds to $n_{ij} = 0$ hereafter.

To give the interviewers incentives to achieve high response rates, Insee provides them with a bonus if they realize the interview. Let $\delta$ denote a bonus and $w$ be a basic wage that Insee might choose. In this case, the interviewer receives for each interview $w + \delta$ when the interview is achieved and $w$ otherwise. Hence, if the interviewer implements a probability $y$ of conducting the survey for each household in his sample, he obtains on average a total wage of $n(\delta y + w)$. We suppose hereafter that interviewers are risk-neutral and have a quasi-linear utility function. In this case, the interviewer $i$, facing a basic wage $w$ and a bonus $\delta$ in survey $j$, will choose a probability $y_{ijn}$ defined for each $n > 0$ by

$$y_{ijn} \in \arg\max_{y} \; n\left[w + \delta y\right] - C_{ijn}(y). \tag{3.1}$$

We now relate the probabilities $y_{ijn}$ chosen by the interviewers with the relevant output, namely the number of interviews $r_{ij}$ interviewer $i$ eventually does in survey $j$. For that, we 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.

**Assumption 1** *(Independence in households reactions) We have, for all $n > 0$,*

$$r_{ij} | n_{ij} = n, (n_{ik})_{k \neq j}, y_{ijn} \sim \mathit{Binomial}(n, y_{ijn}).$$

*When $n_{ij} = 0$, $r_{ij} = 0$.*

Next, we impose some restrictions on the cost function. The first is a separability condition.

**Assumption 2** *(Cost separability)*

$$C_{ijn}(y) = \theta_i C_{jn}(y),$$

*where $C_{jn}$ is twice continuously differentiable, $C'_{jn}(0) = 0$ and $C''_{jn}(y) > 0$ for all $y \in (0, 1)$.*

Assumption 2 imposes that the cost function is separable between an individual-area effect $\theta_i$ 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 & 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 & Tirole, 1993) as well as in empirical works (see Wolak, 1994, Ferrall & Shearer, 1999 or Lavergne & Thomas, 2005). In our application, the term $\theta_i$ captures heterogeneity in the cost of effort across interviewers, but also differences among households from one area to another. For simplicity, we refer subsequently to interviewers' type, but one should keep in mind this dual aspect of $\theta_i$.

Assumption 2 also implies that the probability chosen by the interviewer is defined by the first-order condition of (3.1). In other words, this probability $y_{jn}(\theta)$ satisfies, for all $n > 0$,[8]

$$n\delta_j = \theta C'_{jn}(y_{jn}(\theta)). \tag{3.2}$$

Moreover, differentiating this condition shows that $\theta \mapsto y_{jn}(\theta)$ is strictly decreasing. We denote by $\theta_{jn}(y)$ its inverse function.

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

**Assumption 3** *(No effect and randomness of subsample size) For all $n \in Supp(n_j)$, $C_{jn}(.) = nC_j(.)$. Moreover, $n_j \perp\!\!\!\perp \theta | n_j > 0$.*

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 an household if many interviews have to be done during the day. Assumption 3 is then reasonable if both aspects are of second order, or offset each other. Assumption 3 also imposes that conditional on participation, the number of interviews given to an interviewer is independent of his type.[9] 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

[8] In the absence of ambiguity, we omit the subscript $i$ hereafter.

[9] Because of potential selection effects, we do not impose that $\theta$ is independent of participation ($\mathbb{1}\{n_j > 0\}$).

households to its best interviewers. Because of this concern and since Assumption 3 has actually testable implications, we perform several tests of it, which lend support to its validity (see Appendix A).

### 3.2 Nonparametric identification

We now turn to the empirical content of the interviewer's model. We consider an ideal framework where the number of interviewers in the 2001, 2002 and 2003 surveys is supposed to be infinite. In this case, the distribution function $F_{r_j,n_j}$ of the number of respondents $r_j$ and subsample size $n_j$ of an interviewer can be supposed to be known for both surveys. The question is whether the marginal cost functions $C'_j$ and the distribution of types can be recovered from these functions and the model. To identify these structural functions, we intensively use the assumption that the bonus changes exogenously.

**Assumption 4** *(Exogenous change in the contracts)* $C'_0 = C'_1 = C'_2$.

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 4 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 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. 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 do not differ significantly.

Hence, any change in the cost of interviewers that may have occurred in 2003 should be specific to this particular PCV survey. But the October 2003 PCV survey was drawn in the same way, conducted during the same period and had identical rules for the fieldwork than its predecessors. Hence, there is no obvious reason why the cost should have changed in 2003. The other explanation would be that the acceptance rates has changed because of the topic of the survey. However, the acceptance rates are rather constant over time, around 95% in the PCV surveys (Le Lan, 2008). These rates are very high as these surveys are 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, 2008). 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.

For all 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 an increase in the "social value" of the information in the 2003 survey.[10] The fact 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 indeed have attracted more efficient interviewers. Such a selection effect is compatible with Assumption 4, as this condition remains silent on the selection process of interviewers. We only impose that the same interviewer would have faced the same cost for interviewing people with a given probability in the three surveys. Whether or not this interviewer actually participated to the surveys is a different issue.

Assumption 4 and the fact that the 2001 and 2002 surveys have identical bonuses implies that interviewers choose the same probability on both years. Together with Assumptions 3-1, 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. $j = 1$ will refer hereafter to the 2001-2002 surveys. Our first result is that under our assumptions, the distribution of $y_j (j \in \{1, 2\})$ is identified from the the one of $(n_j, r_j)$. To cope with potential selection effects, we introduce hereafter a participation variable $P$ for interviewers. We let $P = 1$ if the interviewer participates to the 2001 or 2002 (or both) survey only, $P = 2$ if he participates to the 2003 survey only and $P = 3$ if he participates to both the 2003 and either the 2001 or 2002 surveys. We then denote by $F_{jp}$ the cdf of $y_j$ conditional on $P = p$.

**Theorem 3.1** *Suppose that Assumptions 1-4 hold and* $\sup Supp(n_j) = +\infty$ *for* $j \in \{1, 2\}$. *Then, for all* $j$ *and* $p$, $F_{jp}$ *is identified from the distribution of* $(r_j, n_j)$ *conditional on* $P = p$.

The intuition behind this result is that conditional on $n_j = n$, $r_j$ follows a binomial mixture model (see, e.g., Wood, 1999 and for a recent application, D'Haultfoeuille & Rathelot, 2014). In such models, the first $n$ moments of $y_j$ are identified. By independence between $y_j$ and $n_j$ and because $n$ can be chosen arbitrarily large, this implies that all moments of $y_j$ are identified. The result follows because the distribution of a bounded variable is uniquely determined by its moments (see, e.g., Gut, 2005). Note that if $\sup \text{Supp}(n_j) < +\infty$, $F_{jp}$ would be partially identified. But in this binomial mixture model, the problem of partial identification is typically negligible for $\sup \text{Supp}(n_j)$ as small as 6 and sample sizes as large as 10,000 (see D'Haultfoeuille

[10]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.

& Rathelot, 2014). So partial identification of $F_{jp}$ is very likely not a concern in our case where $\max_i n_{i1} = 50$ and $\max_i n_{i2} = 53$ in the data.

Now let us turn to $C'$ and $F_{\theta|p} \equiv F_{\theta|P=p}$. First, a normalization is necessary since for any $\alpha > 0$, we can replace $(\theta, C')$ by $(\alpha\theta, C'/\alpha)$ and leave the model unchanged. Hence, for a given $\theta_0 > 0$, we can choose any $y_0$ in $(0,1)$ such that $\theta_1(y_0) = \theta_0$, where $\theta_1(.)$ denotes the inverse function of $y_1$. Once a normalization has been done on $\theta_1(.)$, no other normalization on $\theta_2$ is needed. This is because the normalization on $\theta_1(y_0)$ induces a normalization on $C'$ (see Equation (3.4) below), which in turn applies to $\theta_2(y_0)$. We also impose the following condition on the distribution of $\theta$.

**Assumption 5** *(Continuous distribution of types) $F_{\theta|p}$ has support $\mathbb{R}^+$ and is continuously differentiable with density $f_{\theta|p}$.*

The support condition imposes that there exists large $\theta$. As explained above, this does not mean that Insee hires very inefficient interviewers, because $\theta$ also reflects the difficulties associated to the area. Considering a support equal to $[a, b]$ with $0 \leq a < b < \infty$ would not modify the analysis below, except that we would have to take care of support issues.[11]

We first focus on stayers here ($P = 3$), for whom there is no selection issue. The idea is to compare $F_{13}$ and $F_{23}$ and use the first-order condition of agents' program to obtain information on $C'$ and $F_{\theta|3}$. In turn, we can use this information to partially recover the distribution of types of the movers $F_{\theta|p}$, for $p \in \{1, 2\}$. We introduce two transforms that are at the basis of our identification method in the presence of an exogenous change. First, a stayer of type $\widetilde{\theta}$ chooses a probability of response equal to $y_1(\widetilde{\theta})$ in survey 1 and $y_2(\widetilde{\theta})$ in survey 2. Because $y_1(\widetilde{\theta})$ and $y_2(\widetilde{\theta})$ are decreasing with $\widetilde{\theta}$, their rank in the distributions $F_{13}$ and $F_{23}$ are identical:

$$\begin{aligned} F_{13}(y_1(\widetilde{\theta})) &= \Pr(y_1(\theta) \leq y_1(\widetilde{\theta}) | P = 3) \\ &= \Pr(\theta \geq \widetilde{\theta} | P = 3) \\ &= \Pr(y_2(\theta) \leq y_2(\widetilde{\theta}) | P = 3) \\ &= F_{23}(y_2(\widetilde{\theta})). \end{aligned}$$

These equations, combined with Assumption 5, also show that $F_{13}$ and $F_{23}$ are strictly increasing. Then, introducing the horizontal transform $H$ defined by $H(y) = F_{23}^{-1} \circ F_{13}(y)$, we get

$$y_2(\widetilde{\theta}) = H(y_1(\widetilde{\theta})). \tag{3.3}$$

[11] Specifically, we would have $\text{Supp}(y_1|P = 3) \neq \text{Supp}(y_2|P = 3)$, implying that the sequence $(\theta_k)_{k\in\mathbb{Z}}$ considered below would be defined only for a subset of $\mathbb{Z}$. But Theorem 3.2 would still hold up to this modification. Note also that in theory, $\sup \text{Supp}(\theta|P = p) + \infty$ is testable, because it implies that $\inf \text{Supp}(y_j|P = p) = 0$. Given that $F_{jp}$ is identified and estimated indirectly through the conditional distribution of $(r_j, n_j)$, testing this condition in practice is however difficult and beyond the scope of the paper.

As the distribution functions $F_{13}$ and $F_{23}$ are identified, $H$ also is. Hence, the knowledge of $y_1(\widetilde{\theta})$ implies the knowledge of $y_2(\widetilde{\theta})$, and the other way round. From an economic perspective, this equality simply states that it is possible to recover the probability an interviewer of type $\widetilde{\theta}$ would choose in survey 2 if we know which probability he chooses in survey 1. To do so, even if his type $\widetilde{\theta}$ is unobserved, it is sufficient to pick the quantile of $F_{23}$ corresponding to $F_{13}(y_1(\widetilde{\theta}))$.

We also rely on the agent's program by using his first-order condition, which satisfies, for an interviewer of type $\theta_j(y)$,

$$\delta_j = \theta_j(y)C'(y). \tag{3.4}$$

Hence, defining the vertical transform $V$ by $V(\theta) = \frac{\delta_1}{\delta_2}\theta$, we obtain, for all $y \in (0, 1)$,

$$\theta_1(y) = V(\theta_2(y)). \tag{3.5}$$

$V$ is identified, so that the knowledge of $\theta_2(y)$ implies the knowledge of $\theta_1(y)$. Contrary to the horizontal transform which links different response rates that similar interviewers choose in both surveys, the vertical transform links interviewers of different types but who choose the same probability in both surveys. Knowing the type of an interviewer with an optimal response rate of $y$ in survey 2, it is possible to recover the type of the interviewer that chooses the same level $y$ in survey 1.

Figure 2 illustrates our identification strategy. We can recover point (1) if we know point (0) through the horizontal transform. Similarly, starting from point (1), we can identify point (2) through the vertical transform. Hence, starting from $(y_0, \theta_0 = \theta_1(y_0))$, we can identify $(y_1, \theta_1 = \theta_1(y_1))$ by $y_1 = H(y_0)$ and $\theta_1 = V(\theta_0) = (\delta_1/\delta_2)\theta_0$. By induction, we identify all the black points in Figure 2. Formally, let $H^k(y) = H \circ ... \circ H(y)$ if $k > 0$, $y$ if $k = 0$ and $H^{-1} \circ ... \circ H^{-1}(y)$ if $k < 0$. Define the sequence $(y_k)_{k\in\mathbb{Z}}$ by $y_k = H^k(y_0)$. Because $y_1(\theta) < y_2(\theta)$ for all $\theta$, $H(y) > y$ for all $y \in (0, 1)$ and $(y_k)_{k\in\mathbb{Z}}$ is increasing.

Then $\theta_k \equiv \theta_1(y_k)$ is simply identified by $\theta_k = \left(\frac{\delta_1}{\delta_2}\right)^k \theta_0$. Elsewhere, $\theta_1(.)$ can be bounded, using the property that it is a decreasing function. Finally, we can use this knowledge on $\theta_1(.)$ to partially identify $C'$ and $F_{\theta|p}$. By the first-order condition (3.4), we simply get

$$C'(y_k) = \frac{\delta_1}{\theta_k}.$$

Elsewhere, $C'$ is partially identified, as it is increasing.

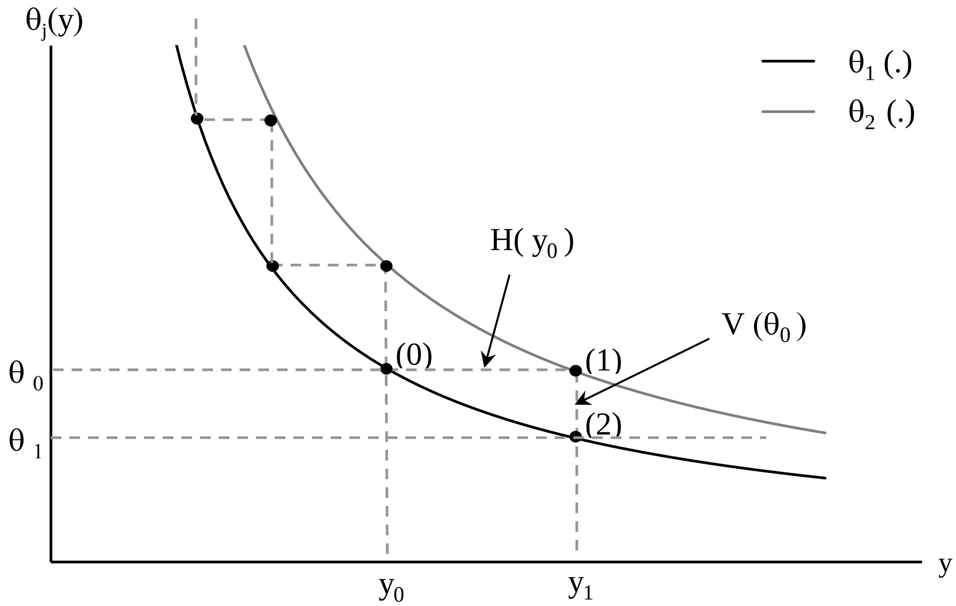



Figure 2: The horizontal and vertical transforms.

Turning to the distribution of $\theta$, we have

$$F_{\theta|p}(\widetilde{\theta}) = 1 - F_{jp}(y_j(\widetilde{\theta})). \tag{3.6}$$

This implies that we point identify $F_{\theta|3}(\theta_k)$ by $1 - F_{13}(y_k)$. Similarly, for the movers,

$$F_{\theta|p}(\theta_k) = \left| \begin{array}{ll} 1 - F_{11}(y_k) & \text{for } p = 1 \\ 1 - F_{22}(y_{k+1}) & \text{for } p = 2. \end{array} \right.$$

We also get bounds outside the sequence $(\theta_k)_{k\in\mathbb{Z}}$ by monotonicity on $F_{\theta|p}$. Theorem 3.2 below provides the exact form of these bounds and ensures that they are sharp.[12]

**Theorem 3.2** *Suppose that Assumptions 2-5 hold and* $\sup Supp(n_j) = +\infty$ *for* $j \in \{1, 2\}$. *Then for all* $y > 0$ *and* $\theta > 0$, $C'(y) \in [\underline{C}'(y), \overline{C}'(y)]$ *and* $F_{\theta|p}(\theta) \in [\underline{F}_{\theta|p}(\theta), \overline{F}_{\theta|p}(\theta)]$, *with*

$$\underline{C}'(y) = \frac{\delta_1}{\theta_{\underline{k}(y)}}, \quad \overline{C}'(y) = \frac{\delta_1}{\theta_{\overline{k}(y)}}, \tag{3.7}$$

$$\underline{F}_{\theta|p}(\theta) = \left| \begin{array}{ll} 1 - F_{1p}(y_{\overline{k}(\theta)}) & \textit{for } p \in \{1, 3\} \\ 1 - F_{22}(y_{\overline{k}(\theta)+1}) & \textit{for } p = 2 \end{array} \right. , \tag{3.8}$$

$$\overline{F}_{\theta|p}(\theta) = \left| \begin{array}{ll} 1 - F_{1p}(y_{\underline{k}(\theta)}) & \textit{for } p \in \{1, 3\} \\ 1 - F_{22}(y_{\underline{k}(\theta)+1}) & \textit{for } p = 2 \end{array} \right. . \tag{3.9}$$

*where* $\underline{k}(y) = \sup\{k \in \mathbb{Z} : y_k \le y\}$, $\overline{k}(y) = \inf\{k \in \mathbb{Z} : y_k \ge y\}$, $\underline{k}(\theta) = \sup\{k \in \mathbb{Z} : \theta_k \ge \theta\}$ *and* $\overline{k}(\theta) = \inf\{k \in \mathbb{Z} : \theta_k \le \theta\}$. *These bounds are identified and sharp. Finally,* $C'$ *and* $F_{\theta|p}$ *are point identified respectively on the sequences* $(y_k)_{k\in\mathbb{Z}}$ *and* $(\theta_k)_{k\in\mathbb{Z}}$.

[12] These bounds are pointwise sharp but not functionally sharp.

Theorem 3.2 provides the best nonparametric bounds on the agents' cost function and the distribution of heterogeneity of interviewers participating to survey $j$. Because Theorem 3.2 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 if 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 $F_\theta$ can be rationalized by the data.

Similar results have been obtained in the auction literature. Guerre et al. (2009) show that exogenous changes are needed to identify first-price auction models with risk averse bidders. A difference between Theorem 3.2 and their result is that we only obtain partial identification here, while in their framework the model is point identified. This is due to the fact that in their case, the bidders' strategies cross at the lowest valuation, and this crossing point can be used for identification. Here $\theta_1(.)$ and $\theta_2(.)$ also cross at $y = 1$, but basically, the crossing point cannot be used for point identification because $\theta_j(1) = 0$. With different type of contracts, the $\theta_j(.)$ could cross inside the support of $y_j$, in which case we would recover point identification (see D'Haultfœuille & Février, 2010).

Finally, our result imply that standard parametric models on $C'$ and $F_{\theta|p}$ are identified with an exogenous change. For instance, the parameters of a lognormal, Weibull or gamma distribution are identified thanks to the knowledge of $F_{\theta|p}$ on the sequence $(\theta_k)_{k\in\mathbb{Z}}$. Actually, because we retrieve an infinite sequence of points on $C'$ and $F_{\theta|p}$, such standard parametric models are overidentified. The sequences $(C'(y_k))_{k\in\mathbb{Z}}$ and $(F_{\theta|p}(\theta_k))_{k\in\mathbb{Z}}$ may thus serve as a guidance for choosing appropriate parametric restrictions, as will be the case in the next section.

### 3.3 Nonparametric estimation of $C'(.)$ and $F_\theta(.)$

We now turn to the estimation of $C'$ and $F_{\theta|p}$. We let $\mathcal{S}$ denote the pooled sample of all interviewers. We study the behavior of our estimators when $N = \min(N_1, N_2, N_3) \to \infty$, where $N_p$ is the number of interviewers with participation status $p$. We impose the following standard assumption of independent sampling.

**Assumption 6** *(independent sampling)* $(\theta_i, r_{i1}, n_{i1}, r_{i2}, n_{i2})_{i\in\mathcal{S}}$ *are i.i.d.*

Our nonparametric estimation method follows closely the identification strategy and may be decomposed into two steps. We first estimate the distributions $F_{jp}$ of the unobserved probabilities $y_1(\theta)$ and $y_2(\theta)$ for interviewers with participation status $p$. We then estimate bounds on the primitive functions $C'$ and $F_{\theta|p}$, using the result of Theorem 3.2.

For the first step, we use a sieve maximum likelihood estimator (see, e.g., Chen, 2006, for a survey on sieve estimation). We choose to approximate the densities[13] $f_{jp}$ by functions of the sieve space

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

where $\mathcal{P}_J$ denotes the space of polynomials of order at most $J$, $M$ is a constant and $(K_N)_{N\in\mathbb{N}}$ is an increasing sequence tending to infinity. We thus approximate the density $f_{jp}$ by squares of polynomials which 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.[14] To see this, let us consider $f(.;\mathbf{a}) \in \mathcal{F}_N$ defined by

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

where $\mathbf{a} = (a_0, ..., a_{K_N})$ and $b_k(\mathbf{a}) = \sum_{l=\max(0,k-K_N)}^{\min(k,K_N)} a_l a_{k-l}$. The likelihood of an observation corresponding to $f(.;\mathbf{a})$ is, by independence between $y_j(\theta)$ and $n_j$ conditional on $P = p$,

$$\begin{aligned}
\Pr(r_j = r | n_j = n, P = p) &= E\left[\Pr(r_j = r | n_j = n, y_j(\theta), P = p)\right] \\
&= \binom{r}{n} E\left[ y_j(\theta)^r (1 - y_j(\theta))^{n-r} | P = p \right] \\
&= \binom{r}{n} \int_0^1 \sum_{k=0}^{2K_N} b_k(\mathbf{a}) y^{r+k} (1-y)^{n-r} dy \\
&= \binom{r}{n} \sum_{k=0}^{2K_N} b_k(\mathbf{a}) B(r+k+1, n-r+1),
\end{aligned}$$

where $B(.,.)$ denotes the beta function. We let $\widehat{f}_{jp}$ denote the maximum likelihood estimator (over $\mathcal{F}_N$) of $f_{jp}$. We then estimate $F_{jp}$ and $F_{jp}^{-1}$ by $\widehat{F}_{jp}(x) = \int_0^x \widehat{f}_{jp}(u)du$ and $\widehat{F_{jp}^{-1}}(u) = \widehat{F}_{jp}^{-1}(x)$.

We now turn to the estimation of $C'$ and $F_{\theta|p}$. First, letting $\widehat{H}(x) = \widehat{F}_{23}^{-1} \circ \widehat{F}_{13}(x)$, we estimate the sequence $(y_k)_{k\in\mathbb{Z}}$ by $\widehat{y}_k = \widehat{H}^k(y_0)$ for all $k \in \mathbb{Z}$. Note that $(\theta_k)_{k\in\mathbb{Z}}$ does not need to be estimated. Then, using Theorem 3.2, the bounds on $C'$ and $F_{\theta|p}$ are estimated by

$$\widehat{\underline{C}'}(y) = \frac{\delta_1}{\theta_{\widehat{\underline{k}}(y)}}, \quad \widehat{\overline{C}'}(y) = \frac{\delta_1}{\theta_{\widehat{\overline{k}}(y)} \theta_0},$$

$$\widehat{\underline{F}}_{\theta|p}(\theta) = \left| \begin{array}{ll} 1 - \widehat{F}_{1p}(\widehat{y}_{\overline{k}(\theta)}) & \text{for } p \in \{1,3\} \\ 1 - \widehat{F}_{22}(\widehat{y}_{\overline{k}(\theta)+1}) & \text{for } p = 2 \end{array} \right. ,$$

[13] Assumption 2 and Equation (3.6) ensure that the densities of $y_1$ and $y_2$ do exist.

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

$$\widehat{\overline{F}}_{\theta|p}(\theta) = \left| \begin{array}{ll} 1 - \widehat{F}_{1p}(\widehat{y}_{\underline{k}(\theta)}) & \text{for } p \in \{1,3\} \\ 1 - \widehat{F}_{22}(\widehat{y}_{\underline{k}(\theta)+1}) & \text{for } p = 2 \end{array} \right. ,$$

where $\underline{k}(\theta)$ and $\overline{k}(\theta)$ are defined as before and $\widehat{\underline{k}}(y) = \sup\{k \in \mathbb{Z} : \widehat{y}_k \leq y\}$ and $\widehat{\overline{k}}(y) = \inf\{k \in \mathbb{Z} : \widehat{y}_k \geq y\}$. To ensure the consistency of our estimators, we impose the following conditions on the cost function and on the distribution of the subsample size $n_j$.

**Assumption 7** *$\lim_{\theta\to\infty} \theta^2 f_{\theta|p}(\theta) = 0$ for $p \in \{1,2,3\}$ and $\lim_{y\to 1} \frac{C''(y)}{C'(y)^2}$ exists and is finite. For all $u > 0$ and $j \in \{1,2\}$, $E(u^{n_j}|P = p) < \infty$ for $p \in \{1,2,3\}$.*

The first condition is very mild and is satisfied for all standard densities with finite expectation. The second condition rules out cases where the function $1/C'(y)$ converges too fast to zero as $y \to 1$. Finally, the third condition imposes light tails for $n_j$. Theorem 3.3 shows that these conditions are sufficient for the consistency of the estimated bounds.

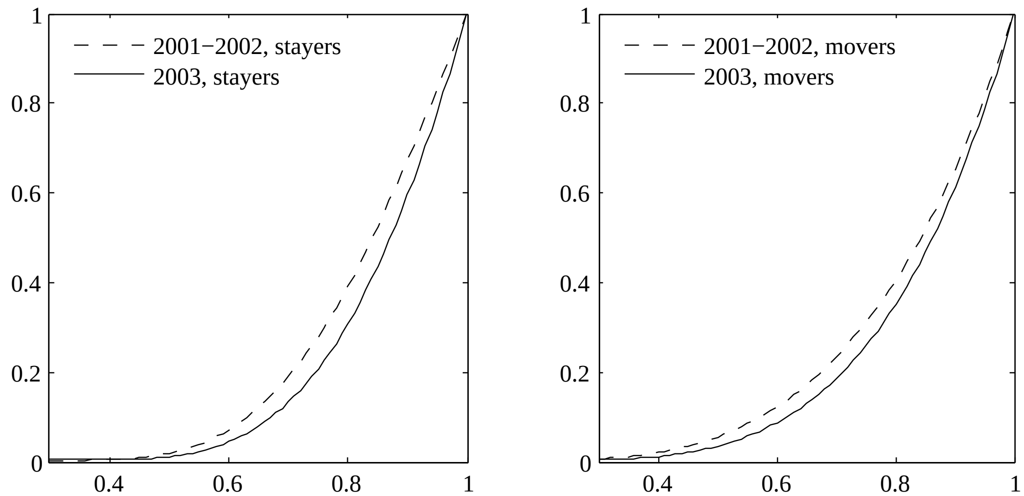
**Theorem 3.3** *Suppose that Assumptions 2, 3, 6 and 7 hold, $K_N \to \infty$ and $K_N^2 \ln K_N/N \to 0$. Then $\widehat{\underline{F}}_{\theta|p}(\theta)$ and $\widehat{\overline{F}}_{\theta|p}(\theta)$ are consistent for $p \in \{1,2,3\}$ and all $\theta > 0$. $\widehat{\underline{C'}}(y)$ and $\widehat{\overline{C'}}(y)$ are consistent on every $y \notin \{y_k, k \in \mathbb{Z}\backslash\{0\}\}$. Moreover, for all $k \in \mathbb{Z}$,*

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

Theorem 3.3 has three parts. The first establishes consistency of the bounds of $F_{\theta|p}$ on its whole support. The second shows the convergence of $\underline{C}'$ and $\overline{C}'$ outside the sequence $(y_k)_{k\in\mathbb{Z}}$. 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 $\left(\widehat{y}_k, \widehat{\overline{C'}}(\widehat{y}_k)\right)$. As a consequence $C'$ and $F_{\theta|p}$ are well estimated on the sequences where they are point identified, while sharp bounds are consistently recovered anywhere else.

### 3.4 Results

We estimate in a first step $F_{jp}$ by the sieve MLE proposed above. As usually, there is a trade-off between bias and variance in the choice of $K_N$. Empirically, the estimates do not seem to be too smooth or too erratic for $K_N$ between 3 and 6. Results are quite similar in this range, and we choose $K_N = 4$ for the stayers and $K_N = 3$ 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.




Notes: 161 observations for $p = 1$, 79 for $p = 2$ and 374 for $p = 3$.

Figure 3: Sieve MLE estimates of $F_{jp}$ for $j \in \{1, 2\}$ and $p \in \{1, 2, 3\}$.

We now estimate nonparametrically the sharp bounds on $F_{\theta|P\in\{2,3\}}$ and $C'$. $F_{\theta|P\in\{2,3\}}$ is interesting as it corresponds to the distribution of interviewers' types on the 2003 survey. We obtain similar patterns for $F_{\theta|p}$, $p \in \{1, 2, 3\}$. We first choose a starting value $y_0$ close to the median of $\widehat{F}_{13}$, namely $y_0 = 0.8$, in order to get more precise estimates for central values of $F_{\theta|P\in\{2,3\}}$ and $C'$.[15] For that $y_0$, we impose the normalization $\theta_1(y_0) = 1$. Figure 4 displays the estimates of the bounds on $F_{\theta|P\in\{2,3\}}$ 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 $F_{\theta|P\in\{2,3\}}$ (resp. $C'$) increases with $|\theta - 1|$ (resp. $|y - 0.8|$), reflecting the fact that, as expected, the estimation error increases with $|n|$.

[15] 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.

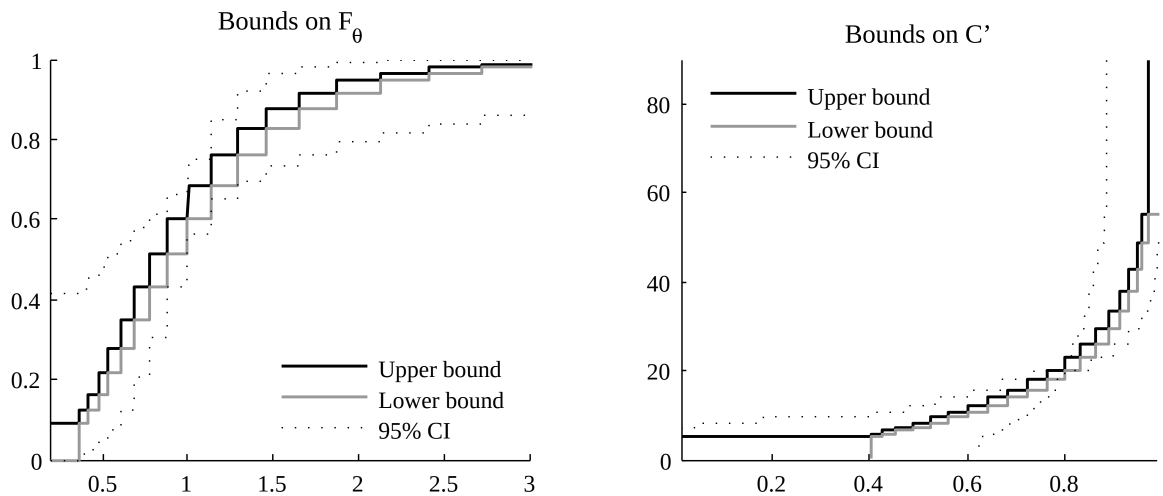



Figure 4: Estimated bounds on $F_{\theta|P\in\{2,3\}}$ and $C'$.

# 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 a contract that differs from the actual ones, we have to take into account selection effects, namely the fact 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 that 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). To perform such a test, remark that $F_{\theta|1} = F_{\theta|2}$ together with (3.6) implies that $F_{21} = F_{22}$. We cannot test directly this condition, because by definition the cdf $F_{21}$ of $y_2(\theta)$ for the 2001-2002 movers is not identified. On the other hand, note that $F_{1p}^{-1} \circ F_{2p} = y_2 \circ \theta_1$ is independent of $p$. This implies that $F_{21} = F_{11} \circ F_{13}^{-1} \circ F_{23}$. Thus, under the null hypothesis that $F_{21} = F_{22}$, we have

$$F_{22} - F_{11} \circ F_{13}^{-1} \circ F_{23} = 0. \tag{4.1}$$

In other words, if we transform the distribution of the probabilities chosen by the 2003 movers using the quantile-quantile transform of the stayers, we should obtain the distribution of the probabilities chosen by the 2002 movers. We consider the test statistic $T = \sup_{y\in[0,1]} |\widehat{\Delta}(y)|$, where $\widehat{\Delta}$ is the nonparametric estimator of

$$\Delta = F_{22} - F_{11} \circ F_{13}^{-1} \circ F_{23}.$$

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 $r_j$ conditional on $P = p$ and $n_j = n$ that satisfy $H_0$ and are consistent under $H_0$:

1. For the stayers and the 2001-2002 movers, we first draw $n_j$ from its empirical distribution, and independently of $n_j$, $y_j$ according to the sieve MLE estimator $\widehat{F}_{jp}$. We then draw $r_j|n_j, y_j \sim \text{Binomial}(n_j, y_j)$.

2. For the 2003 movers, we draw $n_2$ from its empirical distribution, and independently of $n_2$, $y_2$ according to $\widehat{F}_{11} \circ \widehat{F}_{13}^{-1} \circ \widehat{F}_{23}$. We then draw $r_2|n_2, y_2 \sim \text{Binomial}(n_2, y_2)$.

Our estimators of $F_{jp}$ are consistent by Theorem 3.3. Moreover, the bootstrap distribution corresponding to $F_{22}$ satisfies the null hypothesis by construction. Thus, the bootstrap distribution we consider is consistent under the null hypothesis.

We obtain $T \simeq 0.03$ and a p-value of 0.96, and thus cannot reject the absence of selection effects. This result may seem surprising, given the importance of selection effects obtained by, e.g., 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 other optimal contracts

Turning to Insee's program, we suppose that Insee values each interview in survey $j$ as $\lambda_j$. $\lambda_j$ represents the "price" of the information contained in a household's answers. The dependence in $j$ 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 his 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 his 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 aim is to evaluate the loss by Insee due to restricting to linear contracts, keeping in mind that there could 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 $j$ and for a household associated to an interviewer of type $\theta$ is therefore $(\lambda_j - \delta)y_j(\theta, \delta) - w$ when the bonus is $\delta$ 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, $w_j$ is fixed so as to ensure that the worse type participates. Regarding $\delta_j$, we obtain, by optimality of the observed payment scheme among linear contracts and aggregating over all types,

$$\delta_j = \arg\max_{\delta} E\left[(\lambda_j - \delta)y_j(\theta, \delta)\right]. \tag{4.2}$$

This implies that $\delta_j$ satisfies the first-order condition

$$-E\left[y_j(\theta, \delta_j)\right] + (\lambda_j - \delta_j)E\left[\frac{\partial y_j}{\partial \delta}(\theta, \delta_j)\right] = 0. \tag{4.3}$$

Given our assumptions above, this first-order condition can be shown to be both necessary and sufficient.

Given its policy, Insee's expected current surplus for one household in survey $j$ is

$$\Pi_j = (\lambda_j - \delta_j)E\left[y_j(\theta, \delta_j)\right] - w_j. \tag{4.4}$$

This surplus 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 $t^*_{jn}$ for interviewers with

$n$ interviews to conduct takes the form of a vector $(t^*_{jn0}, ..., t^*_{jnn})$, where $t^*_{jnk}$ is the optimal payment of Insee to an interviewer who makes $k$ out of $n$ possible interviews. To define this optimal menu, we first maintain the assumption that Insee ensures universal participation. This implies that $t^*_{jn0} = nw_j$. Second, let $y_{jn}(\theta, t)$ be the probability chosen by an interviewer of type $\theta$ with $n$ households to interview, when facing the contract $t = (t_0, ..., t_n)$. $y_{jn}(\theta, t)$ satisfies

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

The average surplus Insee obtains from an interviewer of type $\theta$, when fixing the transfer vector to $t$ is $\lambda_j y_{jn}(\theta, t)$. The average cost for Insee is

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

Therefore, the optimal menu of contract $t^*_{jn}$ satisfies

$$t^*_{jn} = \arg \max_{t=(t_0,...,t_n) \in \{nw_j\} \times \mathbb{R}^n} E\left[\lambda_j n y_{jn}(\theta, t) - \sum_{k=0}^{n} \binom{n}{k} y_{jn}(\theta, t)^k (1 - y_{jn}(\theta, t))^{n-k} t_k\right]. \tag{4.6}$$

We can also compare the previous surpluses with the one Insee would obtain without asymmetric information, i.e. observing the type of each interviewer and the probability he chooses. Under complete information, Insee is able to fix the probability with which each interviewer will interview his households. These optimal probabilities $y^C_j(\theta)$ satisfy

$$\lambda_j = \theta C'(y^C_j(\theta)). \tag{4.7}$$

Moreover, Insee recovers all the rent from the interviewers. As a result, the optimal transfer function $t^C_j$ with $n$ households to interview is defined by

$$t^C_j(y^C_j(\theta)) = \theta C(y^C_j(\theta)) + nw_j, \tag{4.8}$$

where we have made here the normalization $C(0) = 0$. The expected surplus per household satisfies, under complete information,

$$\Pi^C_j = E\left[\lambda_j y^C_j(\theta) - t^C_j(y^C(\theta))\right]. \tag{4.9}$$

Finally, to analyze further the role of asymmetric information, it is possible to compare these surpluses 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.

### 4.3 Identification and estimation of current and counterfactual surpluses

The current and counterfactual surpluses are all related to $\lambda_j$, so not surprisingly, the issue of their identification boils down to whether $\lambda_j$ 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

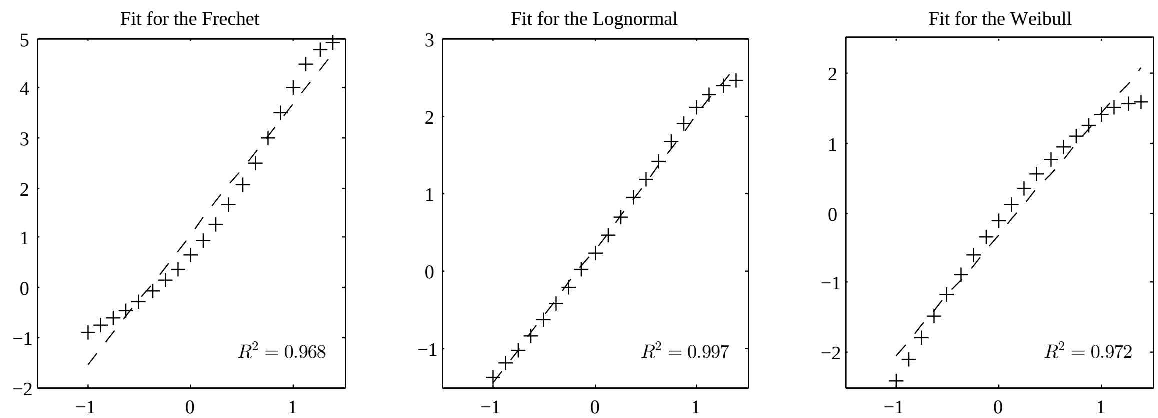
$$\lambda_j = \delta_j + \frac{E[r_j/n_j|P \in \{j,3\}]}{E\left\{1/[\theta C''(y_j(\theta))]|P \in \{j,3\}\right\}}. \qquad (4.10)$$

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

**Proposition 4.1** *Suppose that $C'$ is point identified on $(0,1)$. Then $\lambda_j$ and the parameters defined in (4.4)-(4.9) are identified.*

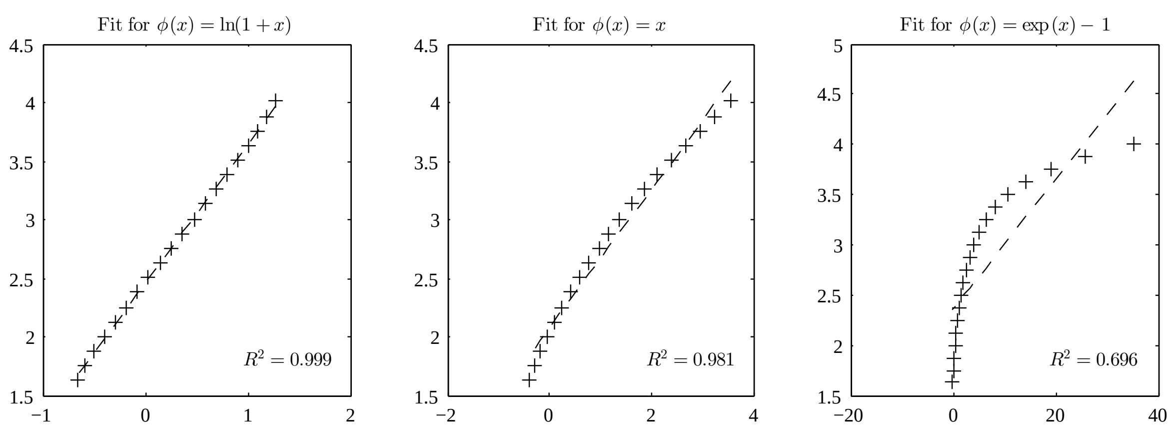
We thus consider a parametric estimation of $(F_{\theta|p})_{p=1,2,3}$ and $C'$. An important aspect is that we use the nonparametric estimates $\left(C'(y_k), F_{\theta|p}(\theta_k)\right)_{k\in\mathbb{Z}}$ to investigate which parametric family fits best. We compare three standard family of distributions on $\mathbb{R}^+$ for $F_{\theta|p}$, namely the Fréchet, for which $F_{\theta|p}(\theta) = \exp(-a\theta^{-b})$ $(a,b>0)$, the lognormal, for which $F_{\theta|p}(\theta) = \Phi((\ln\theta - a)/b)$ (where $\Phi$ denotes the cumulative distribution function of a standard normal variable and $b>0$) and the Weibull, for which $F_{\theta|p}(\theta) = 1-\exp(-a\theta^b)$ $(a,b>0)$. 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). To discriminate between these three families, we plot respectively $-\ln\left(-\ln F_{\theta|p}(\theta_k)\right)$, $\Phi^{-1}\left(F_{\theta|p}(\theta_k)\right)$ and $\ln\left(-\ln(1-F_{\theta|p}(\theta_k))\right)$ 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, $-\ln\left(-\ln F_{\theta|p}(\theta_k)\right) = \ln(a) - b\ln\theta_k$. Similarly, we consider families of marginal cost functions tending to 0 at 0 and to $\infty$ 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 $\phi$ 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, and that these fits are actually very good. The lognormal distribution is also preferred for the movers ($P=1$ and $P=2$). Even if the likelihood ratio tests of nonnested hypotheses (see Vuong, 1989) for the nine corresponding parametric models lead to similar conclusions (the lognormal distribution with $\phi(x) = \ln(1+x)$ or $\phi(x) = x$ being the preferred specifications), it is important to note that such parametric tests only compare models against each others. On the contrary, our procedure allows to test the validity of a parametric family alone, and to choose separately the best parametric family for $C'$ and $F_{\theta|p}$. Thanks to our nonparametric analysis, we do not only learn that the lognormal distribution and $\phi(x) = \ln(1 + x)$ is the best specification among the nine tested, but also that they fit the data correctly.




Notes: plot of $-\ln[-\ln \widehat{F}_{\theta|3}(\theta_k)]$ (left graph), $\Phi^{-1}(\widehat{F}_{\theta|3}(\theta_k))$ (middle graph) and $\ln[-\ln(1 - \widehat{F}_{\theta|3}(\theta_k))]$ (right graph) against $\ln(\theta_k)$. Points should be aligned for the true parametric family. The dotted lines are the best linear approximations.

Figure 5: Choice of the parametric family for $F_{\theta|3}$.




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

Figure 6: Choice of the parametric family for $C'$.

With this parametric specification on $F_\theta$ and $C'$ at hand, we can estimate the model by maximum likelihood. Let $\eta = (\alpha, \beta, a_1, a_2, a_3, b)$. First, observe that under our parametric restriction,

$$y_j(\theta|\eta) = 1 - \exp\left(-\left(\frac{\delta_j}{\alpha\theta}\right)^{1/\beta}\right). \tag{4.11}$$

The probability of observing $(r_1, r_2)$ conditional on $(n_1, n_2)$ then satisfies

$$\begin{aligned}
\Pr(r_1, r_2|n_1, n_2, P = p, \eta) &= \binom{r_1}{n_1}\binom{r_1}{n_1} E\left[y_1(\theta|\eta)^{r_1}(1 - y_1(\theta|\eta))^{n_1 - r_1}\right. \\
&\qquad \left. y_2(\theta|\eta)^{r_2}(1 - y_2(\theta|\eta))^{n_2 - r_2}|P = p\right] \\
&= \binom{r_1}{n_1}\binom{r_2}{n_2} \int y_1(\theta|\eta)^{r_1}(1 - y_1(\theta|\eta))^{n_1 - r_1} \\
&\qquad y_2(\theta|\eta)^{r_2}(1 - y_2(\theta|\eta))^{n_2 - r_2} f_{\theta|p}(\theta|\eta) d\theta,
\end{aligned}$$

and $\eta$ can then be estimated by maximum likelihood on $\mathcal{S}$. Because the likelihood involves integral that do not have closed forms in general, we use simulations hereafter to approximate them. Once we have obtained $\widehat{\eta}$, we use it to derive an estimator of $\lambda_j$ and of all policy parameters, using plug in estimators and the formulas above.

### 4.4 The cost of using inefficient contracts

We first present the maximum likelihood estimates of $\eta$ under the parameter specification chosen above. Results are displayed in Table 3. To take into account possible differences between the stayers and the others, we suppose that $\ln \theta | P = p \sim \mathcal{N}(a_p, b)$. We first estimate the model without constraint. The first column of Table 3 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 significant (p-value=0.33). 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 3). 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.

| 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) |
| $\alpha$ | 11.66 (1.60) | 11.14 (1.68) |
| $\beta$ | 1.16 (0.29) | 1.25 (0.32) |
| $\lambda_1$ | 84.16 (13.75) | 89.31 (15.21) |
| $\lambda_2$ | 103.79 (15.38) | 110.80 (17.01) |

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

Table 3: Maximum likelihood estimates of the parameters of $C'(y) = \alpha\left[\ln(1+y/(1-y))\right]^{\beta}$ and $F_{\theta|p}(\theta) = \Phi\left(\frac{\ln(\theta)-a_p}{b}\right)$.

We now turn to the results on surpluses. We focus on the 2003 survey, the results being very similar for 2001-2002. Table 4 summarizes our results. 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, evaluate the loss of using such simple contracts to be around 50%. Our results point out on the contrary that the cost is quite small and that optimal contracts are not highly nonlinear. This may explain why firms widely use 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 & Sappington (2007), who show that simple tariffs secure at least 89%, 75% and 74% of the maximal surplus, respectively. Studying auctions, Neeman (2003) also proves that simple English auctions generates an expected price that is more than 80% of the value of the object to the bidder with the highest valuation. Finally, studying mixed bundling, Chu et al. (2011) show 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. We also find, in our context, that Insee can use simple contracts and still give the right incentives to its interviewers.

| Environment | Pay method | E[surplus] | Relative | E[response rate] |
|---|---|---|---|---|
| Full information | Optimal contract | 103.04<br>(16.96) | 1.00<br>(0) | 0.99<br>(0.003) |
| Incomplete information | Optimal contract | 80.86<br>(16.95) | 0.78<br>(0.03) | 0.93<br>(0.007) |
| Incomplete information | Linear contract | 68.30<br>(13.86) | 0.66<br>(0.02) | 0.83<br>(0.002) |

Notes: 614 observations. The standard errors were computed using the delta method.

Table 4: Surplus and response rates under alternative compensation schemes.

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' characteristics. We first estimate how the characteristics $X_i$ of interviewer $i$ relate to $\theta_i$, by positing

$$\ln \theta_i = X_i \gamma + \nu_i,$$

where, in line with our lognormal specification, we suppose $\nu_i | X_i \sim \mathcal{N}(0, \sigma^2)$. The characteristics include experience, gender, marriage status, the interviewer's status and the type of area (large urban areas versus others).[16] We reestimate the model keeping our preferred specification of the cost function. The results are displayed in Table 5. 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 7.3%. Note also that the estimators of $\beta$ is very similar to the one we obtained before.

[16] 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.

| Variable | Unrestricted model | Restricted model |
|---|---|---|
| Intercept | $-0.004$ (0.093) | $-0.102$ (0.065) |
| Experience $\leq 5$ | $0.332^{**}$ (0.102) | $0.333^{**}$ (0.097) |
| Experience between 5 and 15 | $0.193^{*}$ (0.077) | $0.186^{**}$ (0.071) |
| Rural or small urban area | $-0.228^{**}$ (0.074) | $-0.212^{**}$ (0.066) |
| Female | 0.002 (0.066) | – |
| Married | $-0.053$ (0.051) | – |
| Stayer | $-0.088$ (0.058) | – |
| $\sigma$ | $0.419^{**}$ (0.106) | $0.398^{**}$ (0.096) |
| $\beta$ | $1.218^{**}$ (0.302) | $1.156^{**}$ (0.27) |
| $R^2 = 1 - \sigma^2/V(\ln\theta)$ | 0.073 | 0.063 |

Notes: 600 and 601 observations are used in the unrestricted and restricted model. A positive parameter indicates larger $\theta$, and thus, on average, lower response rates. The standard errors were computed using the inverse of the estimated hessian, and the delta method. Significance levels: $^{**}1\%$, $^{*}5\%$, $^{\dagger}10\%$.

Table 5: Estimation of the parametric model with interviewers' covariates.

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 5), 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 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 & Shearer (1999), Insee faces social costs due to quite strong unions opposed to such discriminations.

Our results suggest that the gain of using nonlinear contracts would be more substantial, on average around 18.3%, even if our figure is far below the 50% obtained by Ferrall & Shearer

(1999). Even for $n = 2$ households to interview, where the optimal menu of contracts under asymmetric information involves only two parameters as opposed to one with linear contracts, the gain of using this optimal menu compared to the linear one is already of 15.1%. Actually, this gain stabilizes very quickly with $n$: 17.9% for $n = 3$ and then around 18.3% for $n \geq 4$. Because moral hazard disappears as $n \to \infty$ (because then Insee can observe $y(\theta)$ through $r/n$), this shows that the cost of moral hazard here is almost negligible, around 0.0035%.

To understand this, note that moral hazard prevents Insee from considering transfers based on the probability $y(\theta)$ of interview. Only transfers based on $r_2$ (for any fixed $n$) are possible. But because interviewers are risk-neutral and $\widetilde{t}_n(y) \equiv E[t_{r_2}/n_2|n_2 = n, y]$ are polynomials of $y$ of order $n$, the situation is as if Insee could define contracts based on $y(\theta)$ but was constrained to use polynomials of order $n$ of $y(\theta)$ instead of considering any function of $y(\theta)$. This will not be too much an issue if the optimal contract can be well approximated by a polynomial. Figure 7 plots the optimal 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 almost negligible loss, as the number of households for which $n = 2$ only represent 0.14% of the whole sample of households.

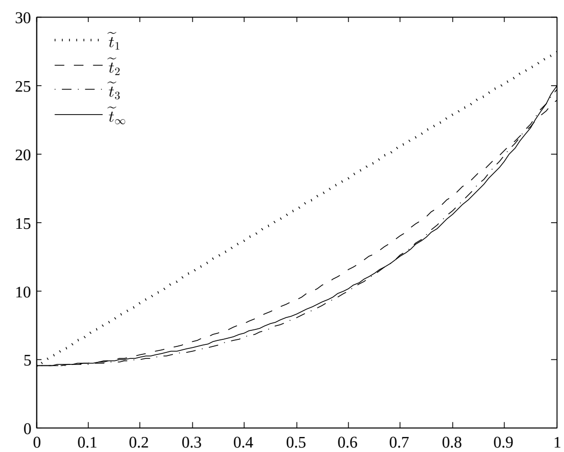



Figure 7: Evolution of $\widetilde{t}_n = E[t_{r_2}/n_2|n_2 = n, y]$ with $n$.

Hence, adverse selection, that is to say not observing or not using interviewers' type, appears to be really the issue, more than moral hazard. Note that if using optimal asymmetric contracts would significantly increase Insee's surplus, it may be difficult for Insee to implement such contracts. Indeed, our results imply that only 4.2% of the interviewers (the less efficient ones) would benefit from such a change.

## 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.

## References


Aryal, G. (2014), An empirical analysis of competitive nonlinear pricing. Working paper.

Aryal, G., Perrigne, I. & Vuong, Q. (2010), Identification of insurance models with multidimensional screening. Working paper.

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

Campanelli, P. C., Martin, E. A. & Rothgeb, J. M. (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.

Chen, X. (2006), Large sample sieve estimation of semi-nonparametric models, *in* J. J. Heckman & E. E. Leamer, eds, 'Handbook of Econometrics', Vol. 6, North Holland.

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

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

Chu, L. Y. & Sappington, D. E. M. (2007), 'Simple cost-sharing contracts', *The American Economic Review* **97**, 419–428.

Chung, D., Steenburgh, T. J. & Sudhir, K. (2013), Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. Harvard Business School Working Paper 13-066.

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

Copeland, A. & Monnet, C. (2009), 'The welfare effect of incentive schemes', *Review of Economic Studies* **76**, 93–113.

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

D'Haultfœuille, X. & Février, P. (2010), Identification of a class of adverse selection models with contracts variation. Mimeo.

D'Haultfœuille, X. & Février, P. (2015), 'Identification of nonseparable triangular models with discrete instruments', *Econometrica* **83**, 1199–1210.

D'Haultfoeuille, X. & Rathelot, R. (2014), Measuring segregation on small units: A partial identification analysis. Working paper.

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

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

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

Guerre, E., Perrigne, I. & Vuong, Q. (2009), 'Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions', *Econometrica* **77**, 1193–1227.

Gut, A. (2005), *Probability: A Graduate Course*, Springer-Verlag, New York.

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

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

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

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

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

Lazear, E. (2000), 'Performance pay and productivity', *American Economic Review* **90**, 1346–1361.

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

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

Luo, Y., Perrigne, I. & Vuong, Q. (2014), Structural analysis of nonlinear pricing. Working paper.

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

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

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

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

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

O'Muircheartaigh, C. & Campanelli, P. (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.

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

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

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

Perrigne, I. & Vuong, Q. (2010), Nonlinear pricing in yellow pages. Mimeo.

Perrigne, I. & Vuong, Q. (2011), 'Nonparametric identification of a contract model with adverse selection and moral hazard', *Econometrica* **79**, 1499–1539.

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

Raju, J. S. & Srinivasan, V. (1996), 'Quota-based compensation plans for multiterritory heterogeneous salesforces', *Management Science* **6**, 1454–1462.

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.

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

Torgovitsky, A. (2015), 'Identification of nonseparable models using instruments with small support', *Econometrica* **83**, 1185–1197.

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

Vuong, Q. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**, 307–333.

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

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

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

## 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 6 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 informations 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 three-days 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 & Kandel, 1988 or O'Muircheartaigh & Campanelli, 1998). It seems, however, that experienced interviewers are less prone to this so-called interviewer's effet (see, e.g., Cleary et al., 1981, Singer et al., 1983 or Campanelli et al., 1991). Finally, most 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 6 also shows that the typical interviewer is a middle-aged woman who is out of the labour 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 4,545 euros earned on average by household interviewers in 2001 corresponds to the minimum wage for a third time job.

| Variable | All interviewers in 2001 | PCV interviewers 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 | 4,054 (3,075) | 6,146 (3,139) | 2,976 (1,702) | 5,169 (2,300) |
| 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 |

Notes: 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.

Table 6: Descriptive statistics on Insee household interviewers.

## B Tests of Assumption 3

Assumption 3 implies that the targeted probability of interviewers is not related to the size of the subsample they have to handle. Such a condition is testable. We consider a first test based on the fact that under Assumption 3, we have, for $j \in \{0, 1, 2\}$,

$$E(r_j/n_j|n_j, P) = E(y_j(\theta)|n_j, P) = E[y_j(\theta)|P]. \quad \text{(B.1)}$$

This implies that $r_j/n_j$ is mean independent of $n_j$ conditional on $P$. Once we control for participation, the observed response rate $r_j/n_j$ should not depend on expectation on $n_j$ because the probabilities chosen by interviewers do not depend on $n_j$ ($C_{jn} = nC_j$) and because interviewers' type do not depend on $n_j$. We test the restriction (B.1) by considering the model

$$r_{ij}/n_{ij} = \zeta_{jP_i} + g(n_{ij}) + \varepsilon_{ij},$$

where $E(\varepsilon_{ij}|n_{ij}, P_i) = 0$ and $\zeta_{jp}$ is a year times participation status fixed effect. Equation (B.1) implies that the function $g$ should be equal to zero. We perform a test of this restriction using linear, quadratic and a 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 $r_j/n_j$ and $n_j$ under the alternative.[17]

| 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 $\times$ 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 |

Notes: 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 $\theta_i$. Significance levels: $^{**}1\%, ^{*}5\%, ^{\dagger}10\%$.

Table 7: Test of Assumption 3 based on regressions of response rates on functions $g$ of subsample sizes.

Results are presented in Table 7. 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.

The first test uses both the fact that $y_{jn}(.)$ does not depend on $n$, which stems from the condition $C_{jn}(\theta) = nC_j(\theta)$, and independence between $\theta$ and $n$. Our second and third tests aim to check separately these two conditions, relying on parametric forms that are consistent with our nonparametric estimates (see Subsection 3.4). First, suppose that $C'_{jn}(y) = \alpha_j f(n)\left(\frac{y}{1-y}\right)^\beta$. Then the first-order condition yields

$$y_{jn}(\theta) = \frac{1}{1 + \exp\left(-\frac{1}{\beta}\left[\ln(\delta_j/\alpha_j) - \ln\theta - \ln(f(n)/n)\right]\right)}.$$

Combined with Assumption 1, this implies that the dummy $y_{ijk}$ of whether interviewer $i$ managed or not to interview household $k$ in survey $j$ satisfies

$$y_{ijk} = \mathbb{1}\{\gamma_j + \widetilde{f}(n_{ij}) + \widetilde{\theta}_i + \varepsilon_{ijk} \geq 0\}, \tag{B.2}$$

[17] Our results are robust to other flexible specifications.

where $\gamma_j = \ln(\delta_j/\alpha_j)/\beta$, $\widetilde{f}(n) = -[\ln f(n)/n]/\beta$, $\widetilde{\theta}_i = -\frac{1}{\beta}\ln(\alpha\theta_i)$ and the $(\varepsilon_{ijk})_{i,j,k}$ are independent and follow a logit distribution. We can therefore test whether $C'_{jn} = nC'_j$, without imposing conditional independence between $n_{ij}$ and $\theta_i$, by estimating a conditional logit model on the $(y_{ijk})_{i,j,k}$ of the stayers, and testing $\widetilde{f} = 0$. The idea is that the fixed effect controls for potential dependence between $\theta_i$ and $n_{ij}$, while $\widetilde{f}$ reflects the effect of $n_{ij}$ on $y_{ijk}$ through the dependence in $n$ of $C'_{jn}/n$.

Second, under the additional restriction that $\widetilde{\theta}$ is normal, which is our preferred parametric restriction hereafter (see Subsection 4.3), we can also test for independence between $n_j$ and $\theta$ conditional on $P$, without imposing $C'_{jn} = nC'_j$. It suffices indeed to test whether the estimator obtained by a random coefficient model differs from the one obtained by the fixed effect conditional logit model, through a Hausman test. Because we cannot estimate the conditional logit on movers, we actually test here $n_j \perp\!\!\!\perp \theta | P = 3$.

| Variable | Linear $\widetilde{f}$ | Quadratic $\widetilde{f}$ | Piecewise linear $\widetilde{f}$ |
|---|---|---|---|
| Fixed effect estimation | | | |
| Subsample size | $-0.0102$ (0.0077) | $-0.0171$ (0.0191) | $-0.023$ (0.0217) |
| Square of the subsample size | - | 0.0002 (0.0004) | - |
| (Subsample size – 10)$^+$ | - | - | 0.0176 (0.0277) |
| (Subsample size – 20)$^+$ | - | - | $-0.006$ (0.0222) |
| p-value of the test $\widetilde{f} = 0$ (or, equivalently, $C'_{jn} = nC'_j$) | 0.186 | 0.386 | 0.539 |
| Random effect estimation | | | |
| Subsample size | $-0.0038$ (0.0059) | $-0.0111$ (0.0158) | $-0.0033$ (0.0199) |
| Square of the subsample size | - | 0.0002 (0.0004) | - |
| (Subsample size – 10)$^+$ | - | - | $-0.0037$ (0.0262) |
| (Subsample size – 20)$^+$ | - | - | 0.0061 (0.0193) |
| p-value of the Hausman test $n_j \perp\!\!\!\perp \theta \| P = 3$ | 0.194 | 0.418 | 0.070 |

Notes: estimates of (B.2) with fixed and random (normal) effects logit models. 10,461 observations. Significance levels: $^{**}$1%,$^{*}$ 5%,$^{\dagger}$ 10%.

Table 8: Tests of $C'_{jn} = nC'_j$ and $n_j \perp\!\!\!\perp \theta | P = 3$ under parametric restrictions on $C'_{jn}$ and $F_{\theta|3}$ .

Table 8 displays the result of these two tests. We accept the null hypotheses of both tests at standard levels, despite the relatively large number of observations. Overall, these tests suggest that Assumption 3 is valid. It implies that the response rate chosen by interviewers of type $\theta$ does not depend on $n$, so we denote it by $y_j(\theta)$ hereafter, while its inverse is $\theta_j(y)$.

## C Proofs

### Proof of Theorem 3.1

First, by Assumption 3 and the first-order condition, $y_j(\theta)$ does not depend on $n$. We denote it by $y_j(\theta)$. Then, for all $n$ in the support of $n_j$ and all $1 \leq r \leq n$, we have

$$\begin{aligned}
\Pr(r_j = r | n_j = n, P = p) &= E\left[\Pr(r_j = r | n_j = n, y_j(\theta)) | n_j = n, P = p\right] \\
&= E\left[\binom{n}{r} y_j(\theta)^r (1 - y_j(\theta))^{n-r} | n_j = n, P = p\right] \\
&= E\left[\binom{n}{r} y_j(\theta)^r (1 - y_j(\theta))^{n-r} | P = p\right] \\
&= \sum_{i'=0}^{n-r} \binom{n}{r}\binom{n-r}{i'} (-1)^{n-r-i'} E(y_j(\theta)^{n-i'} | P = p) \\
&= \sum_{i=r}^{n} \binom{n}{i}\binom{i}{r} (-1)^{i-r} E(y_j(\theta)^{i} | P = p) \\
&= \sum_{i=1}^{n} \binom{n}{i}\binom{i}{r} (-1)^{i-r} E(y_j(\theta)^{i} | P = p),
\end{aligned}$$

where the first equality follows from the law of iterated expectation, the second from Assumption 1, the third stems from independence between $\theta$ and $n_j$ conditional on $P = p$ (Assumption 3), the fourth from the decomposition of $(1 - y_j(\theta))^{n-r}$, 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 $\mathbf{P}_n = (\Pr(r_j = 1 | n_j = n, P = p), ..., \Pr(r_j = n | n_j = n, P = p))'$, $m_n = (E(y_j(\theta)^1 | P = p), ..., E(y_j(\theta)^n) | P = p)'$ 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

$$\mathbf{P}_n = \mathbf{Q}_n m_n. \tag{C.1}$$

Moreover, $\mathbf{Q}_n$ is invertible as an upper triangular matrix with non-zero diagonal elements. Thus, $m_n$ is identified by $\mathbf{Q}_n^{-1}\mathbf{P}_n$. Conditional on $P = p$, the $n$ first moments of $y_j(\theta)$ are identified from the distribution of $r_j$ conditional on $n_j = n$. Because $\sup\{n : \Pr(n_j = n | P = p) > 0\} = +\infty$, all moments of $y_j(\theta)$ are identified. This, together with $y_j(\theta)$ bounded, ensures that the distribution of $y_j(\theta)$ conditional on $P = p$ is identified (see, e.g., Gut, 2005). ■

**Proof of Theorem 3.2**

It follows from the discussion before Theorem 3.2 that $\theta_1$ is point identified on $(y_k)_{k\in\mathbb{Z}}$. For other $y$, we get, by monotonicity of $\theta_1$,

$$\theta_{\overline{k}(y)} = \sup_{k\in\mathbb{Z}:\, y_k\geq y} \theta_1(y_k) \leq \theta_1(y) \leq \inf_{k\in\mathbb{Z}:\, y_k\leq y} \theta_1(y_k) = \theta_{\underline{k}(y)}.$$

Similarly,

$$y_{\underline{k}(\theta)} = \sup_{k\in\mathbb{Z}:\, \theta_k\geq \theta} y_k \leq y_1(\theta) \leq \inf_{k\in\mathbb{Z}:\, \theta_k\leq\theta} y_k = y_{\overline{k}(\theta)}.$$

By Equations (3.4) and (3.6), Inequalities (3.7) and (3.8) hold. The last point of the theorem follows directly from the definitions of the bounds on $\theta_1(y)$ and $y_1(\theta)$.

We now show that for all $y \in (0,1)\backslash\{y_k : k\in\mathbb{Z}\}$, and $\theta\in\mathbb{R}^+\backslash\{0,\theta_1(y_k) : k\in\mathbb{Z}\}$, the bounds on $C'(y)$ and $F_{\theta|p}(\theta)$ are sharp. We focus, for a given $y$, on $\overline{C}'(y)$ as the proof is similar for $\underline{C}'(y)$, $\overline{F}_{\theta|p}(\theta)$ and $\underline{F}_{\theta|p}(\theta)$. More precisely, we want to construct a function $\widetilde{C}'$ such that $\widetilde{C}'(y)$ is arbitrarily close to $\overline{C}'(y)$, and which satisfies all the restrictions given by the data and the model.

The proof is in two step. First, fixing $\varepsilon>0$, we construct a continuously differentiable function $\widetilde{\theta}_1$ that satisfies $\widetilde{\theta}_1(y_k)=\theta_1(y_k)$ for all $k\in\mathbb{Z}$ and $\widetilde{\theta}_1(y)=\delta_1/(\overline{C}'(y)-\varepsilon)$. In a second step, we study the function $\widetilde{C}'=\delta_1/\widetilde{\theta}_1$.

For the first step, letting $k\in\mathbb{Z}$ denote the integer such that $y_k<y<y_{k+1}$, we first define $\widetilde{\theta}_1$ on $[y_k, y_{k+1})$. To do so, we consider any strictly decreasing continuously differentiable function $\widetilde{\theta}_1$ such that $\widetilde{\theta}_1(y_k)=\theta_k$, $\widetilde{\theta}_1(y)=\delta_1/(\overline{C}'(y)-\varepsilon)$ and $\lim_{y\to y_{k+1}^-}\widetilde{\theta}_1(y)=\theta_{k+1}$. Moreover, we impose that

$$\lim_{y\to y_{k+1}^-} \widetilde{\theta}'_1(y) = \frac{\delta_1\widetilde{\theta}'_1(y_k)}{\delta_2 H'(y_k)}. \tag{C.2}$$

Such a function exists. We then extend it on $(0,1)$ through the vertical and horizontal transforms. For instance, $\widetilde{\theta}_1$ is defined on $[y_{k+1}, y_{k+2}) = [H(y_k), H(y_{k+1}))$ by

$$\widetilde{\theta}_1(y) = \frac{\delta_1}{\delta_2}\widetilde{\theta}_1(H^{-1}(y)).$$

Moreover, because $H$ is continuously differentiable, $\widetilde{\theta}_1$ admits a right derivative at $y_{k+1}$ given by

$$\lim_{y\to y_{k+1}^+} \widetilde{\theta}'_1(y) = \frac{\delta_1\widetilde{\theta}'_1(y_k)}{\delta_2 H'(y_k)},$$

and Equation (C.2) ensures that $\widetilde{\theta}_1$ is differentiable at $y_{k+1}$. By induction, using either $H$ or $H^{-1}$, it is possible to extend $\widetilde{\theta}_1$ on $(0,1)$ to obtain a continuously differentiable function on the whole interval. This function will also be strictly decreasing as both $H$ or $H^{-1}$ are increasing.

We now consider the function $\widetilde{C}'(y) \equiv \delta_1/\widetilde{\theta}_1(y)$. By construction, the first-order condition $\widetilde{\theta}_1(y)\widetilde{C}'(y) = \delta_1$ and the equality $\widetilde{C}'(y) = \overline{C}'(y) - \varepsilon$ are satisfied. That $\widetilde{C}'$ is strictly positive and strictly increasing follows from its definition and the fact that $\widetilde{\theta}_1$ is strictly decreasing. $\widetilde{C}'$ is also continuously differentiable as $\widetilde{\theta}_1$ is. Finally, by definition of $\widetilde{C}'$,

$$0 = \widetilde{\theta}_1'(y)\widetilde{C}'(y) + \widetilde{\theta}_1(y)\widetilde{C}''(y).$$

Because $\widetilde{\theta}_1'(y)\widetilde{C}'(y) < 0$, we get

$$-\widetilde{\theta}_1(y)\widetilde{C}''(y) < 0,$$

and the second order condition is satisfied. ■

### Non-identification with one menu of contracts

With only one menu of contracts, the status variable $P$ is irrelevant, so we drop it here. Let us consider a strictly increasing and differentiable function $\widetilde{C}'$, different from the true one $C'$. Define then $\widetilde{\theta}_1$ by

$$\widetilde{\theta}_1(y) = \frac{\delta_1}{\widetilde{C}'(y)}.$$

$\widetilde{\theta}_1$ is strictly decreasing and admits an inverse function $\widetilde{y}$. Then define $\widetilde{F}_{\theta|n}$ by

$$\widetilde{F}_\theta(\theta) = 1 - F_1(\widetilde{y}(\theta)).$$

By construction $\widetilde{C}'$ and $\widetilde{F}_\theta$ are consistent with the first and second order conditions and the identified distribution $F_1$. As a result, $C'$ and $F_\theta$ are not identified. ■

### Proof of Theorem 3.3

The proof proceeds in four steps. We first prove that $\widehat{F}_{jp}$ is uniformly consistent. We then prove that $\widehat{H}$ is uniformly consistent on each compact set included in $(0,1)$. Thirdly, we prove that for all $k \in \mathbb{Z}$, $\widehat{y}_k$ is consistent. Finally, we show that the estimated bounds of $C'$ and $F_{\theta|p}$ are consistent.

**1. Uniform consistency of $\widehat{F}_{jp}$.**

For any continuous function $g$ on $[0,1]$ let $\|g\| = \sup_{x\in[0,1]} |g(x)|$. We actually prove the stronger result that for all $(j,p)$ such that $f_{jp}$ is well defined,

$$\left\| \widehat{f}_{jp} - f_{jp} \right\| \xrightarrow{\mathbb{P}} 0. \tag{C.3}$$

First, note that for all $y \in (0,1)$, $f_{jp}(y) = \theta'_j(y) f_{\theta|p}(\theta_j(y))$, so that $f_{jp}$ is continuous on $(0,1)$. Moreover, differentiating the first-order condition, we obtain

$$\theta'_j(y) = -\frac{\theta_j(y)C''(y)}{C'(y)} = -\frac{\delta_j C''(y)}{C'^2(y)}. \tag{C.4}$$

Thus, by Assumption 7, $\lim_{y\to 1} f_{jp}(y)$ exists and is finite. The same holds at 0. Thus, we can extend $f_{jp}$ by continuity 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 $r \in \{0, ..., n\}$, let

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

let $Q_{jp}(f) = E(l(f, r_j, n_j)|P = p)$ denote the conditional expectation of $l(f,r,n)$ with respect to $(r_j, n_j)$ and

$$Q_{N_j,jp}(f) = \frac{1}{N_p}\sum_{i:P_i=p} l(f, r_{ij}, n_{ij}),$$

By definition of $\widehat{f}_{jp}$, $\widehat{f}_{jp} = \arg\max_{f\in\mathcal{F}_N} Q_{N_p,jp}(f)$ is a sieve M-estimator. We use Theorem 3.1 of Chen (2006) and its associated Remark 3.2 to prove (C.3). To this end, we check the following conditions:

a. $Q_{jp}$ is uniquely maximized at $f_{jp}$ and $Q_{jp}(f_{jp}) > -\infty$.

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

c. $Q_{jp}$ is continuous for $\|.\|$.

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

e. $E\left[\sup_{f\in\mathcal{F}_N} |l(f, r_j, n_j)| \,|\, P = p\right] < \infty$.

f. There exists $U(.,.)$ such that $E(U(r_j, n_j)|P = p) < \infty$ and for all $(f,g) \in \mathcal{F}_N^2$, $|l(f, r_j, n_j) - l(g, r_j, n_j)| \leq \|f - g\| U(r_j, n_j)$.

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

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

$$\begin{aligned}
E\left[\frac{\exp l(g, r_j, n_j)}{\exp l(f_{jp}, r_j, n_j)}\bigg| n_j = n, P = p\right] &= \sum_{r=0}^{n} \Pr(r_j = r|n, P = p)\frac{\binom{r}{n}\int_0^1 y^r(1-y)^{n-r}g(y)dy}{\Pr(r_j = r|n, P = p)} \\
&= \int_0^1 \left(\sum_{r=0}^{n}\binom{r}{n} y^r(1-y)^{n-r}\right) g(y)dy \\
&= \int_0^1 g(y)dy \\
&= 1.
\end{aligned}$$

Thus,

$$E\left[\frac{\exp l(g, r_j, n_j)}{\exp l(f_{jp}, r_j, n_j)}\bigg| P = p\right] = 1.$$

Besides, because $f_{jp}$ is identified, we have $l(g, r_j, n_j) \neq l(f_{jp}, r_j, n_j)$ with a strictly positive probability for all $g \neq f_{jpj}$. Thus, by Jensen's inequality,

$$E\left[\ln\left(\frac{\exp l(g, r_j, n_j)}{\exp l(f_{jp}, r_j, n_j)}\right)\middle| P = p\right] < \ln E\left[\frac{\exp l(g, r_j, n_j)}{\exp l(f_{jp}, r_j, n_j)}\middle| P = p\right] = 0.$$

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

$$\begin{aligned}
\int_0^1 y^r(1-y)^{n-r} f_{jp}(y)dy &\geq \int_{u_1}^{1-u_1} y^r(1-y)^{n-r} f_{jp}(y)dy \\
&\geq u_1^n \int_{u_1}^{1-u_1} \left(\frac{y}{u_1}\right)^r \left(\frac{1-y}{u_1}\right)^{n-r} f_{jp}(y)dy \\
&\geq u_1^n \int_{u_1}^{1-u_1} f_{jp}(y)dy \\
&\geq \frac{u_1^n}{2}. \qquad \text{(C.5)}
\end{aligned}$$

As a result, $Q_{jp}(f_{jp}) \geq E(n_j|P = p)\ln u_1 - \ln 2$. By Assumption 7, $E(n_j|P = p) < \infty$, so that $Q_{jp}(f_{jp}) > -\infty$.

b. First, $\mathcal{F}_N \subset \mathcal{F}_{N+1}$ for all $N$ since $K_N$ 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 $\left\|\sqrt{f} - P\right\| \leq \varepsilon$. Then,

$$\begin{aligned}
\|f - P^2\| &\leq \left\|\sqrt{f} - P\right\| \times \left\|\sqrt{f} + P\right\| \\
&\leq \left\|\sqrt{f} - P\right\| \times \left(2\left\|\sqrt{f}\right\| + \left\|P - \sqrt{f}\right\|\right) \\
&\leq \varepsilon\left(\varepsilon + 2\left\|\sqrt{f}\right\|\right).
\end{aligned}$$

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

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

We have

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

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

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

so that $f_N \in \mathcal{F}_N$. Moreover,

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

This establishes b, since the right-hand side tends to zero with $\varepsilon$.

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 $r \in \{0, ..., n\}$,

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

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

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

As a result,

$$|Q_{jp}(f) - Q_{jp}(g)| \leq E\left[|l(f, r_j, n_j) - l(g, r_j, n_j)|\right] \leq 2\varepsilon E\left(\frac{1}{u_2^n}|P = p\right).$$

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

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

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

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

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

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

Thus,

$$E\left[\sup_{f\in\mathcal{F}_N} |l(f,r_j,n_j)|\right| P=p\Bigg] \leq E\left[|\ln m(n_j,r_j)|\ |P=p\right] \leq E\left[|\ln 2| + n|\ln u_3|\ |P=p\right],$$

and $E(n|P=p) < \infty$ implies that $E\left[\sup_{f\in\mathcal{F}_N} |l(f,r_j,n_j)|\ |P=p\right] < \infty$.

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

$$|l(f,r_j,n_j) - l(g,r_j,n_j)| \leq \frac{2\,\|f-g\|}{u_3^{n_j}}.$$

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

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

$$\left\|f'\right\| \leq 2(2K_N)^2\,\|f\| \leq 8MK_N^2 \ln K_N.$$

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

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

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

$$\ln N_b(\delta, \mathcal{G}_N, \|.\|) \leq C_0 K_N^2 \ln K_N.$$

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

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

We now establish that for all $0 < \underline{x} < \overline{x} < 1$,

$$\sup_{x\in[\underline{x},\overline{x}]} |\widehat{H}(x) - H(x)| \xrightarrow{\mathbb{P}} 0.$$

In view of $H(x) = F_{23}^{-1} \circ F_{13}(x)$, we first prove that for any compact $K$ strictly included in $(0,1)$,

$$\sup_{y\in K} |\widehat{F}_{23}^{-1}(y) - F_{23}^{-1}(y)| \xrightarrow{\mathbb{P}} 0 \tag{C.10}$$

By Assumption 2 and 5, $\theta_2'(y) < 0$ and $f_\theta(\theta_2(y)) > 0$ for all $y \in (0,1)$. Hence, by continuity of $f_{\theta|3}$ and $\theta_2'$, for all compact $K$ included in $(0,1)$,

$$\min_{y\in K} f_{23}(y) = \min_{y\in K} \left[-f_{\theta|P=3}(\theta_2(y))\theta_2'(y)\right] > 0. \tag{C.11}$$

If $\varepsilon > 0$ is such that $E = \{x \in \mathbb{R} : \exists y \in F_{23}^{-1}(K) : |x-y| \leq \varepsilon\}$ is a subset of $(0,1)$, (C.11) implies that $C_1 = \min_{y\in E} f_{23}(y) > 0$. Moreover, by the mean value theorem, for all $y \in K$,

$$F_{23}(F_{23}^{-1}(y) - \varepsilon) + C_1\varepsilon \leq F_{23}(F_{23}^{-1}(y)) \leq F_{23}(F_{23}^{-1}(y) + \varepsilon) - C_1\varepsilon.$$

Consequently,

$$\begin{aligned}
& \Pr\left(\sup_{y\in K} |\widehat{F}_{23}^{-1}(y) - F_{23}^{-1}(y)| > \varepsilon\right) \\
= \ & \Pr\left(\exists y \in K : \widehat{F}_{23}^{-1}(y) > F_{23}^{-1}(y) + \varepsilon \cup \widehat{F}_{23}^{-1}(y) < F_{23}^{-1}(y) - \varepsilon\right) \\
= \ & \Pr\left(\exists y \in K : \widehat{F}_{23}(\widehat{F}_{23}^{-1}(y)) = F_{23}(F_{23}^{-1}(y)) > \widehat{F}_{23}(F_{23}^{-1}(y)+\varepsilon) \cup F_{23}(F_{23}^{-1}(y)) < \widehat{F}_{23}(F_{23}^{-1}(y)-\varepsilon)\right) \\
\leq \ & \Pr\left(\exists y \in K : F_{23}(F_{23}^{-1}(y)+\varepsilon) - \widehat{F}_{23}(F_{23}^{-1}(y)+\varepsilon) > C_1\varepsilon \text{ or } \widehat{F}_{23}(F_{23}^{-1}(y)-\varepsilon) - F_{23}(F_{23}^{-1}(y)-\varepsilon) > C_1\varepsilon\right) \\
\leq \ & \Pr\left(\left\|\widehat{F}_{23} - F_{23}\right\| > C_1\varepsilon\right).
\end{aligned}$$

Because $\widehat{F}_{23}$ converges uniformly to $F_{23}$, (C.10) holds.

Now, fix $\varepsilon > 0$ and $\zeta > 0$ such that $F_{13}(\underline{x}) > \zeta$ and $F_{13}(\overline{x}) < 1-\zeta$. For all $N$ large enough,

$$\Pr\left(\left\|\widehat{F}_{13} - F_{13}\right\| > \zeta\right) \leq \varepsilon/2. \tag{C.12}$$

If $\left\|\widehat{F}_{13} - F_{13}\right\| \leq \zeta$, we get, for all $x \in [\underline{x}, \overline{x}]$, noting $K = [F_{13}(\underline{x}) - \zeta, F_{13}(\overline{x}) + \zeta]$,

$$\begin{aligned}
|\widehat{H}(x) - H(x)| &\leq |\widehat{F}_{23}^{-1}(\widehat{F}_{13}(x)) - F_{23}^{-1}(\widehat{F}_{13}(x))| + |F_{23}^{-1}(\widehat{F}_{13}(x)) - F_{23}^{-1}(F_{13}(x))| \\
&\leq \sup_{u\in K} |\widehat{F}_{23}^{-1}(u) - F_{23}^{-1}(u)| + C_2\left\|\widehat{F}_{13} - F_{13}\right\|,
\end{aligned} \tag{C.13}$$

where $C_2 = \sup_{u\in K} F_{23}^{-1\prime}(u) < \infty$ by (C.11). Fix $\delta > 0$. By uniform convergence of $\widehat{F}_{23}$ and (C.10), for all $N$ large enough,

$$\Pr\left(\sup_{u\in K} |\widehat{F}_{23}^{-1}(u) - F_{23}^{-1}(u)| + C_2\left\|\widehat{F}_{13} - F_{13}\right\| > \delta\right) < \frac{\varepsilon}{2}. \tag{C.14}$$

Then, for all $N$ large enough,

$$\begin{aligned}
\Pr\left(\sup_{x\in[\underline{x},\overline{x}]} |\widehat{H}(x) - H(x)| > \delta\right) &\leq \Pr\left(\sup_{x\in[\underline{x},\overline{x}]} |\widehat{H}(x) - H(x)| > \delta, \left\|\widehat{F}_{13} - F_{13}\right\| \leq \zeta\right) \\
&\quad + \Pr\left(\left\|\widehat{F}_{13} - F_{13}\right\| > \zeta\right) \\
&\leq \Pr\left(\sup_{u\in K} |\widehat{F}_{23}^{-1}(u) - F_{23}^{-1}(u)| + C_2\left\|\widehat{F}_{13} - F_{13}\right\| > \delta\right) + \frac{\varepsilon}{2} \\
&\leq \varepsilon,
\end{aligned}$$

where the second inequality stems from (C.12) and (C.13), and the third from (C.14). The result follows since $\varepsilon$ and $\delta$ were arbitrary.

**3. Consistency of $\widehat{y}_k$, for all $k \in \mathbb{Z}$.**

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

$$\Pr(|\widehat{y}_k - y_k| \leq \varepsilon) \to 1 \tag{C.15}$$

Let us proceed by induction on $k$. The proposition is true when $k = 0$. Suppose that it holds for $k - 1 \geq 0$ and let us prove that it holds for $k$ (the proof is similar for negative values). By definition of $y_k$ and $\widehat{y}_k$, it suffices to prove that for all $\varepsilon > 0$,

$$\Pr(|\widehat{H}(\widehat{y}_{k-1}) - H(y_{k-1})| \leq \varepsilon) \to 1 \tag{C.16}$$

Without loss of generality, we can focus only on $\varepsilon > 0$ such that $B(y_{k-1}, \varepsilon) \subset (0, 1)$, where $B(x, r)$ is the closed ball of center $x$ and radius $r$. Because

$$H'(x) = \frac{f_{23}(x)}{f_{13}\left(F_{23}(x)\right)},$$

it follows, by (C.11), that $C_3 = 1 \vee \sup_{x \in B(y_{k-1}, \varepsilon)} |H'(x)| < \infty$. Moreover, by the induction hypothesis and the uniform convergence of $\widehat{H}$, for all $N$ large enough, the event

$$E_0 = \left\{ |\widehat{y}_{k-1} - y_{k-1}| < \varepsilon/2C_3, \sup_{x \in B(y_{k-1}, \varepsilon)} |\widehat{H}(x) - H(x)| < \varepsilon/2 \right\}$$

holds with an arbitrarily large probability. Under $E_0$,

$$\begin{aligned}
|\widehat{H}(\widehat{y}_{k-1}) - H(y_{k-1})| &\leq |\widehat{H}(\widehat{y}_{k-1}) - H(\widehat{y}_{k-1})| + |H(\widehat{y}_{k-1}) - H(y_{k-1})| \\
&\leq \sup_{x \in B(y_{k-1}, \varepsilon)} |\widehat{H}(x) - H(x)| + C_3 |\widehat{y}_{k-1} - y_{k-1}| \\
&\leq \varepsilon.
\end{aligned}$$

This proves (C.16) and concludes the induction step. Thus, (C.15) holds for all $k \in \mathbb{Z}$.

**4. Consistency of the estimators of the bounds of $C'$ and $F_{\theta|p}$.**

Consistency of the estimators of the bounds of $F_{\theta|p}$ follows from their definitions, consistency of $\widehat{y}_k$ for all $k \in \mathbb{Z}$ and uniform consistency of $\widehat{F}_{jp}$. To prove consistency of $\widehat{\overline{C}'}$ and $\widehat{\underline{C}'}$, it suffices to show that $\widehat{\underline{k}}(y)$ and $\widehat{\overline{k}}(y)$ are consistent for all $y \notin \{y_k, k \in \mathbb{Z}\}$. For such $y$, we have $y_{\underline{k}(y)} < y < y_{\overline{k}(y)}$. Let us consider the event $E_1 = \left\{ \widehat{y}_{\underline{k}(y)} < y < \widehat{y}_{\overline{k}(y)} \right\}$. By convergence of $\widehat{y}_{\underline{k}(y)}$ and $\widehat{y}_{\overline{k}(y)}$, $\Pr(E_1) \to 1$. This proves the result, since we also have $E_1 = \{\widehat{\underline{k}}(y) = \underline{k}(y), \widehat{\overline{k}}(y) = \overline{k}(y)\}$ by definition of $\widehat{\underline{k}}(y)$ and $\widehat{\overline{k}}(y)$. ■

**Proof of Proposition 4.1**

If $C'$ is identified, $\lambda_j$ can be recovered with (4.10). The current surplus can then be obtained directly using Equation (4.4). Similarly, under Assumption 2, $y_{jn}(\theta, t)$ defined by (4.5) only depends on $C'$ and on the distribution of $\theta$ 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 $y_j^C(\theta)$ fixed under complete information are identified from (4.7), as long as $C'$ is fully identified. Because we have supposed that Insee fixes $w_j$ so as to ensure universal participation, We can recover the optimal transfer function $t_{jn}^C$ from Equation (4.8) and $t_{jn}^C(0) = nw_j$ . Identification of $\Pi_j^C$ finally follows from (4.9). ■