Centre de Recherche en Économie et Statistique

CREST

Série des Documents de Travail

n° 2011-29

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

X. D'HAULTFOEUILLE[1]
Ph. FÉVRIER[2]

1 CREST. Email : xavier.dhaultfoeuille@ensae.fr
2 CREST. Email : philippe.fevrier@ensae.fr

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

Xavier D'Haultfoeuille †   Philippe Février‡

May 2011

Abstract

This paper addresses the issue of workers' reaction to incentives, the optimality of simple linear compensation contracts and the importance of asymmetries between firms and workers. We study linear 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 then used to select correctly our parametric restrictions. We find that interviewers react to incentives, their productivity increasing by 5.6% when the piece rate increases by 13.4%. We also show 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. Finally, we find moderate costs of asymmetric information in our data, the loss being around 20% of what Insee could achieve under complete information.

Keywords: Incentives, Asymmetric Information, Optimal Contracts, Nonparametric Identification.

JEL classification numbers: C14, D82, D86

---

*We are very grateful to Steve Berry, Thierry Magnac, Arnaud Maurel, 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 also thank Daniel Verger and the Unité Méthodes Statistiques of Insee for providing us with the data.

†CREST (INSEE). 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., 2009 or Copeland & Monnet, 2009). Our paper adds to this empirical personnel literature by addressing the three following issues: how much do workers react to incentives? Are simple linear compensation contracts nearly optimal? How important are asymmetries between firms and workers?

To answer these questions, we use 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 depend on two factors. First, they 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, and that this change was exogenous for the interviewers. As detailed in the paper, we have reasons to believe that the bonus increase in 2003 stems only from a change in Insee's objective function. In 2003, the focus of the survey was put on education practices in the family, a topical issue at that time and for which Insee needed to get more precise results than in 2001 and 2002. In that respect, the change in the bonus is not related to the interviewers and can be considered as an instrument that affects the contract without being directly related to the cost of interviewers' effort or their heterogeneity. Following this contract change, the productivity of the interviewers increased by 5.6%, the response rates going from 78.7% in 2001 and 2002 to 83.1% in 2003. As the change is exogenous, only selection and incentive effects may explain this finding. More efficient interviewers may be attracted by higher wages (the selection effect), while, at the same time, the bonus increase may push up interviewers' effort and productivity (the incentive effect). Thanks to the panel structure of our data, we are able to disentangle both effects by comparing interviewers hired in all three surveys (the "stayers") with the "movers" hired only in one of them. Contrary to Lazear (2000) who estimates the selection effect to explain half of the productivity increase, we do not find any selection effect in our data. The observed change in the response rate is, in our application, entirely due to incentive effects. The productivity increasing by 5.6% when the bonus increases by 13.4%, we thus estimate an elasticity to piece rates around 0.4. These results are in line with the previous literature. To increase their productivity, firms provide incentives to their workers through compensation contracts, the main instruments being piece rates and bonuses. Both types of contracts appear to work well in practice. Facing such incentives, workers produce higher efforts and increase their productivity. Lazear (2000), for instance, estimates that the productivity increases by more than 20% for workers in the car glass industry when introducing a simple piece rate scheme. Similarly, in a dynamic framework, Copeland & Monnet (2009) find a 12% productivity gain in the check-clearing industry when changing the payment scheme.[1]

To address the two remaining questions, namely the efficiency of simple linear compensation contracts and the importance of asymmetries between firms and workers, we rely on a structural approach. The main difference with the previous personnel literature is that we study, in the first place, the nonparametric identification of our model. We do so in a spirit close to what has been done in the structural auction literature, building on the work of Guerre et al. (2000). We are able to partially identify the cost function and the distribution

[1] See also Paarsch & Shearer (2000), Shearer (2004) and recent results in the marketing literature from, e.g., Steenburgh (2008), Misra & Nair (2008) or Chung et al. (2009).

of the interviewers' types. More precisely, we develop a new induction technique that allows us to point identify both functions of interest on a sequence of points. Elsewhere, we derive sharp bounds on these functions, using a monotonicity argument.[2] An important feature of our identification procedure is that the information on the functions of interest are recovered using the interviewers' program solely. This is convenient because we have reasons to believe that Insee does not implement the optimal contracts but only optimizes over linear ones. Beyond identification, we also develop a nonparametric estimation procedure using our recursive 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 for the interviewers and a parametric specification for the objective function of Insee. As the model is not point identified nonparametrically, such restrictions are necessary to estimate the 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. This issue is important to investigate the interviewers' behavior, the optimality of contracts or to do policy exercises, because the results are sensitive to the parametric choices.

It is also worth noting that this identification method has a broader set of applications. As explained by D'Haultfœuille & Février (2010), it applies to many adverse selection models. The empirical literature on such models has grown rapidly in recent years. Examples include auction models (see e.g. Paarsch, 1992 and Guerre et al., 2000), regulatory contracts (see, among others, Wolak, 1994, Gagnepain & Ivaldi, 2002, Perrigne, 2002, Perrigne & Vuong, 2004 and Lavergne & Thomas, 2005) or nonlinear pricing / price discrimination models (see Ivaldi & Martimort, 1994, Miravete, 2002, Leslie, 2004, Miravete & Roller, 2005, Crawford & Shum, 2007, Huang et al., 2007 and Miravete, 2007). All these models share a common underlying structure for which our procedure is well adapted and can be useful to study their nonparametric identification and estimation.

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

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

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

implementation (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 a recent empirical paper by Miravete (2007) which reports a loss of only 3%. Ferrall & Shearer (1999), on the other hand, concludes that simple nonlinear compensation plans lead to substantial inefficiencies.

Then, our method 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 79% of the full information surplus. This loss (21%) 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 (62%) is due to incomplete information whereas the last 38% are associated with the simple payment scheme.

As already mentioned, our method also allows us to select a parametric specification in line with our nonparametric results. To test the importance of the information recovered in the nonparametric step, we consider several parametric families for the cost function of the interviewers and for the distribution of their types. Depending on the specification, the expected surplus under incomplete information varies between 65% and 83% of the full information surplus, whereas the loss of using a simple linear compensation plan lies between 1% and 16%. The results are thus quite sensitive to the parametric choices. It highlights the importance of having an exogenous change to recover nonparametrically some information in a first step and to select appropriate parametric restrictions based on this information in a second one.

The paper is organized as follows. Section 2 presents institutional details and the data at our disposal. Section 3 develops the theoretical model of the interviewer and Insee. Section 4 focuses on the identification and estimation of the model. The results are displayed in Section 5, and Section 6 concludes. All proofs are deferred to the appendix.

2 Institutional details and data description

2.1 Institutional details

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 do so, Insee draws, approximately every ten years, a large sample of housings[4] from the exhaustive census database.[5] This sample consists of geographical areas called primary units. All survey samples are then drawn from these primary units. Given the sampling structure, Insee hires interviewers for a long period taking into account, among other things, the distance between the primary units and the interviewer's own address. Hence, even if the precise set of interviewers may vary from one survey to another, Insee usually relies on the same pool of interviewers for all its surveys.

There are at least three reasons for this policy. First, 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. Second, 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 1 shows that, as a result of this organization, the average experience of interviewers is 6.8 years at the beginning of 2001. Moreover, out

[4]As in many countries, Insee draws samples of housings rather than of households as it only has an exhaustive database of housings at its disposal.

[5]The introduction of an annual census in 2004 has modified the way this large sample is constituted. Apart from that, the rest of our description still applies today.

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.

Table 1 also displays some socio-demographic characteristics for interviewers hired by Insee in 2001 in these 12 surveys. 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, the annual income of 4095 euros earned on average by the interviewers corresponds to the minimum wage for a third time job.

Variable	Average	Std dev	Min	Max
Experience at Insee (in years)	6.8	8.2	0	42
Number of surveys done in 2001	5.2	3.6	1	12
Income in 2001 euros	4,095	3,262	71	21,119
Woman	80.5%	0.40	0	1
Married	54.8%	0.50	0	1
Other professional activity (Yes=0, No=1)	64.3%	0.48	0	1
Age	43.1	11.7	18	77

Source : Insee

Table 1: Descriptive statistics on Insee interviewers in 2001.

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

2.2 Data

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. 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 2 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[6] (see Figure 1), with a p-value of the one-sided Kolmogorov-Smirnov test equal to 0.003. We also note that the distribution functions displayed in Figure 1 exhibit several jumps, especially at 0.5, 0.67 and 1. These jumps are due to the fact that the response rates are ratios of two integers, and the number of

[6] The average response rate on 2001-2002 is defined as the ratio between the total number of interviews and the total number of households, where the 2001 and 2002 data are pooled.

households to inverview is rather small.[7]

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%

Table 2: Descriptive statistics on the full sample.

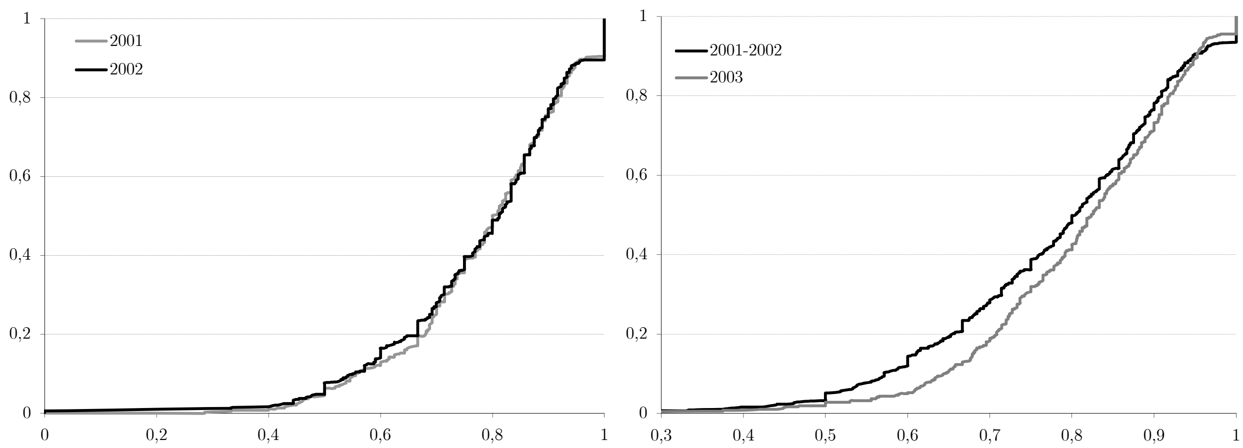

Figure 1: Distribution functions of the response rates on all interviewers, for all households.

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.

[7]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.

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.

Table 3 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 2. 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 3: 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),[8] 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. To summarize, interviewers were paid respectively 25, 24.8 and 27.5 euros for each successful interview and 4.7, 4.6 and 4.6 euros for an unsuccessful one. As interviewers achieved higher response rates in the 2003 survey for which the bonus was higher, incentive effects may be at stake. However, the 2003 increase may also stem from other changes in the survey or from the interviewers themselves. We now explain the reasons underlying this change and why we believe it to be exogenous.

[8] All figures are in 2002 euros.

2.3 An exogenous change

Insee is a public institute whose surveys are analyzed by researchers and used in policy debates. Surveys may thus differ in the "social value" of the information that can be recovered from it. In our case, we believe that the change in 2003 stems from a modification of these values. In 2001 and 2002, the focus of the survey was put respectively on the use of new technologies and participation in associations, while in 2003, the survey studied education practices in the family. 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. Given these elements, it is possible that the social value of an interview was higher in 2003, which may explain why Insee decided to increase the bonus and to double the size of the sample. Insee needed the number of respondents to be high in order to get more precise results on this important topic.

Related to this, Insee might have modified its bonus because of the change in the sample size. This would be the case for instance if it were harder to achieve a given response rate for larger sample sizes. To investigate this issue, we regress the response rate z_{ij} for an interviewer i in the survey j on the number of households n_{ij} assigned to interviewer i, controlling for interviewer and survey fixed effects:

$$z_{ij} = \beta n_{ij} + u_i + v_j + \varepsilon_{ij}. \tag{2.1}$$

Within estimates are presented in Table 4. We find that the coefficient β is not significantly different from 0 at a 5% level, which indicates that there is no effect of the sample size on interviewers' response rate. The coefficient is actually negative, indicating that there might be some economies of scale in interviewers' work. According to our estimates, these economies of scale seem nevertheless to be very small. This is not surprising as housings, even at the interviewer level, can be quite far away from each other in these surveys. Consequently, the change in the sample size can not explain the higher response rate observed in 2003. It only affects the accuracy of the estimates obtained from this survey.

Coefficients	Estimate	
Constant	0.80**	(0.013)
Subsample size	-0.0022	(0.0013)
Year 2002	0.11	(0.11)
Year 2003	0.61**	(0.15)

Significativity levels: **1%,* 5%.

Table 4: Effect of the subsample size on response rates.

One might also suspect that good interviewers receive more households to survey, in order for the Insee to increase the total number of respondents. In this case, the number of households interviewers receive would be correlated with their fixed effect, so that

$$n_{ij} = \gamma u_i + v'_j + \eta_{ij},$$

where v'_j is a survey fixed effect different from v_j. Fixing β to zero in (2.1) as it is not significant, and replacing u_i by its expression, we obtain

$$n_{ij} = \gamma z_{ij} + w_j + (\eta_{ij} - \gamma\varepsilon_{ij}),$$

where $w_j = v'_j - \gamma v_j$. z_{ij} is endogenous in this equation because of its correlation with ε_{ij}, but we can instrument it by z_{ij-1}. The results in Table 5 indicate that there is no relationship between the subsample size and productivity of the interviewers. This result is reassuring. Indeed, as explained previously, the sample is drawn at the national level and each interviewer receives the sample that corresponds to his geographic area. Our result suggests that Insee is limited by these geographical constraints and cannot allocate freely the households to its interviewers. It is defined exogenously by the draw of the sample and the location of the corresponding households.[9]

[9] This point is not directly related to the exogeneity of the contract change in 2003, but still will be important for identification (see our Assumption 2 below).

Coefficients	Estimate
Constant	2.95 (2.51)
Subsample size	5.76 (3.07)
Year 2003	7.76^{**} (0.49)

Significativity levels: **1%,* 5%.

Table 5: Relationship between the subsample size and interviewers' fixed effects.

The bonus could also have changed because the sample was different in 2003. For example, Insee may have increased its bonus if households were known ex ante to be harder to contact. However, as explained previously, the three subsamples of families we consider are drawn in the same way. Hence, it cannot explain any change in the payment scheme.

The observed change may also be related to the interviewers themselves. First, any global shock on the interviewers market may explain the observed increase. This could have been the case, for instance, if, because of a decrease of unemployment, the outside options of interviewers had increased substantially in 2003. Nevertheless, if such effects were at play, the bonus of other 2003 surveys would also have been affected in a similar way. As we do not observe any increase in the bonus of other 2003 surveys, such an explanation is implausible.[10]

As previously explained, the time spent to try to contact the households either by phone or by coming to their house represents the main part of interviewer's cost. Any change in this cost might explain the bonus increase. However, the surveys were drawn in the same way, conducted during the same period and had identical rules for the fieldwork. There is thus no obvious reason why this cost should have changed from one year to another. The only 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

[10]For instance, the compensation schemes of the two regular surveys (namely the labor force survey and the survey on rents and service charges) which took place at the same time were not modified.

in the participation decision. This is reinforced by the fact that the questionnaires of PCV surveys contains a fixed part, always identical for all October editions, which represents more than half of the questions.

Finally, an indirect test of the assumption than nothing changes for the interviewers from one survey to another is to compare the outcome in 2001 and 2002. The payment rules were similar, as well as the way the surveys were drawn and conducted. The fact that the response rates were very similar (see Figure 1) should thus be seen as an evidence that, for a given payment rule, nothing modifies the response rates and the interviewers behavior from one year to another.

For all these reasons, we believe that the 2003 change is exogenous in the sense that it came exclusively from a change in the objective function of Insee, independently of the interviewers.[11] This does not mean, however, that incentive effects entirely explain the pattern observed in Figure 1. The 2003 compensation scheme may indeed have attracted more efficient interviewers, inducing a so-called selection effect. To separate both effects, we compare the interviewers that participated in all three surveys (the "stayers" subsequently) with those who participate in only one survey (the "movers"). Table 6 displays the average response rates for both movers and stayers. Actually, in 2003, stayers obtained an average response rate slightly above the one of the movers (83.8% versus 83.1%), a result which is not compatible with an interviewer selection effect. We were indeed expecting new interviewers to be more productive. Actually, stayers perform slightly better in all surveys (79.6% versus 78.5% in 2001, 80.4% versus 79.6% in 2002), probably reflecting positive returns to experience in this job, although differences in average response rates between stayers and movers are not significant at 5%. As a result, the previous conclusion still applies when restricting ourselves to stayers. As depicted in Figure 2 and formally tested by Kolmogorov-Smirnov tests, the distribution functions on stayers do not differ significantly at 5% in 2001 and 2002, while the 2003 one still stochastically dominates the one of 2001-2002. Hence, contrary to Lazear (2000) who estimates the selection effect to explain half of the productivity increase in his application, we find that the observed change in the response rate is entirely due to incentive effects. 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

[11]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 this time.

effects could only occur through a reallocation of interviewers among this pool. Our result suggests that such reallocations are not related to interviewers' productivity.[12]

Year	Number of Interviewers	Number of movers	Average response rate Stayers	Movers	T-test of the difference
2001	377	137	79.6%	78.5%	0.41
2002	471	101	80.4%	76.6%	0.51
2003	453	79	83.8%	80.4%	0.72

Table 6: Comparison between stayers and movers.

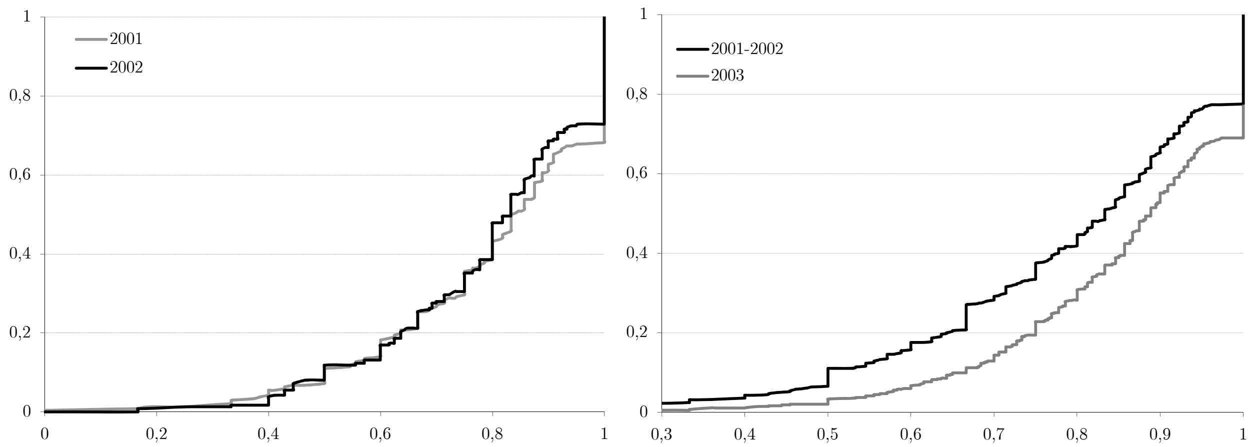

Figure 2: Distribution functions of the response rates on stayers.

2.4 Incentive effects

To sum up, our results strongly suggest that the change in the response rate observed in 2003 comes from the reactions of interviewers to the bonus increase, and from no other reason. It can thus be considered a pure incentive effect. We find that, when increasing the bonus by 13.4% (from 20.2 to 22.9 euros), the response rates increase by 5.6%, going from 78.7% in 2001-2002 to 83.1% in 2003. This effect is similar when restricting ourselves to the stayers, with an increase of 5.1%, from 79.8% to 83.8%. Interviewers have thus reacted

[12]It is nevertheless important to note that, as explained in D'Haultfœuille & Février (2010), the method developed in Section 4 still applies in the presence of selection effects.

to the change in the payment scheme, with an elasticity to the bonus around 0.4.

These results contribute to the personnel literature, which has repeatedly put forwards the positive effects of incentives. Using variations in piece rates given to tree planters in British Columbia, Paarsch & Shearer (1999) actually report larger elasticities, ranging from 0.77 to 2.14. Rather than computing such elasticities, other papers focus on the comparison between piece rates and fixed wages. Using respectively a structural approach and a field experiment, Paarsch & Shearer (2000) and Shearer (2004) show that the productivity of tree planters is around 20% higher with piece rate than with fixed wages. This result is of similar magnitude as those of Lazear (2000) and Copeland & Monnet (2009) on workers in car glass and check-clearing industries, respectively. More generally, in the marketing literature, Steenburgh (2008), Misra & Nair (2008) and Chung et al. (2009) highlight the idea that both the shape and the timing of the compensation schemes matter for firms. Chung et al. (2009), for instance, show that annual bonuses should be combined with quarterly bonuses to increase their impact on productivity.

Figure 3 also displays the density and cumulative distribution function of the difference between response rates in 2001-2002 and in 2003 for stayers. Most of the stayers (57.6%) achieve a better response rate in 2003. However, consistently with the literature, we observe an important heterogeneity in workers' reactions. While the first decile of interviewers encounters a decrease of more than 15%, the last one displays an increase of more than 23%.

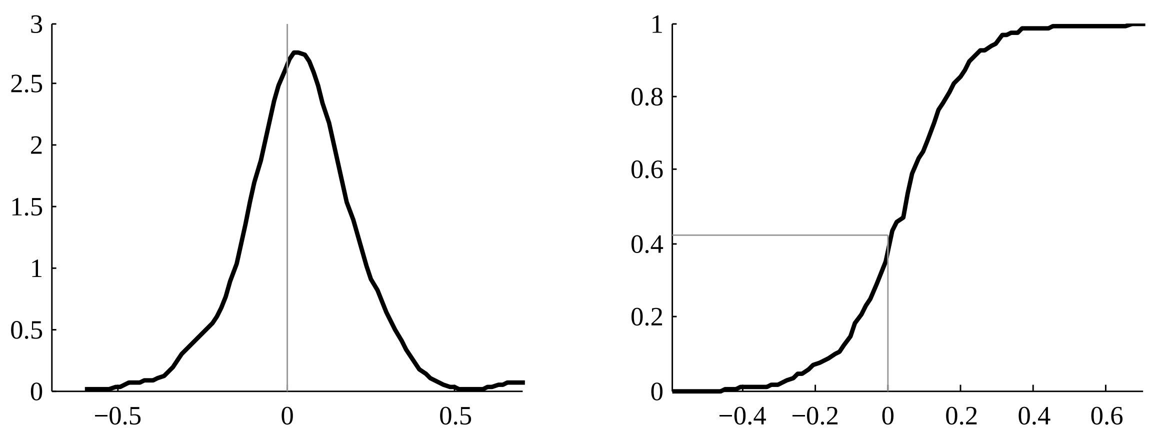

Figure 3: Density and cumulative distribution function of the difference in 2001-2002 and 2003 response rates on stayers

3 The Model

To analyze further the interviewers' behavior and the issue of contracts optimality, we need to rely on a structural approach. We first model interviewers' decision and then turn to Insee's program.

3.1 The interviewers' program

At an individual level, 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. 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 z_{ij} of interviewer i in survey j on the mean of the 1999 census characteristics X_{ij} of his sample. More precisely, we regress the response rates on the mean size of the household, the mean number of rooms of the housing and the mean age of the reference person in the samples, controlling for interviewers and years fixed effects:

$$z_{ij} = X_{ij}\beta + u_i + v_j + \varepsilon_{ij}.$$

Table 7, Column I (resp. Column II), presents the results on the whole sample (resp. on families). In both cases, the 2003 fixed effect is significantly positive whereas the 2002 fixed effect is not significantly different from zero. As previously noticed, higher response rates correspond to higher bonuses. Column I also shows that the response rates increase with the size of the households. This reflects the idea that families are easier to contact. However, when restricting ourselves to families, none of the census variables are significantly different from zero anymore. 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.

Coefficients	Column I	Column II
Constant	0.71^{**} (0.046)	0.67^{**} (0.077)
Household size	0.027^{*} (0.013)	0.018 (0.017)
Number of rooms	0.003 (0.012)	0.011 (0.012)
Age of the reference person	-0.0002 (0.0002)	0.0000 (0.0001)
Year 2002	-0.008 (0.007)	0.012 (0.011)
Year 2003	0.018^{*} (0.008)	0.044^{**} (0.012)

Table 7: Fixed effect linear regression of response rates on average housing characteristics at the 1999 census.

To sum up, a given interviewer exogenously receives a sample of homogenous families that he has to survey. We have also seen (see Table 4 above) that the size of the subsample does not play any role in his decision. These results support the idea that he decides which effort to exert household by household. Because families are homogenous in terms of contact ease, he treats them similarly and takes the same decision for all of them. Heterogeneity in the response rates achieved by different interviewers only arises because of intrinsic differences between them or their designated area.

An interviewer has thus to decide, for each household, with which probability y he wants to survey each of his household and has to produce his effort accordingly. The probability y thus corresponds to the response rate an interviewer wants to achieve on his sample.[13] The expectation of the cost for interviewer i to obtain a probability y in survey j is denoted by $C_{ij}(y)$. It represents the expected cost to contact an household and convince him to accept the interview with probability y. As mentioned in Subsection 2.3, there is no obvious reason why this cost should have changed from one year to another, once we restrict ourselves to families. The interviewers population was very similar in the three surveys, as well as the sampling procedure. The surveys were drawn in the same way, conducted during the same period and had identical rules for the fieldwork. The change in the bonus was exogenous and not related with any change on the interviewers' side. We thus suppose that the cost C_{ij} does not depend on j. We summarize the heterogeneity of the interviewers and their associated area by a parameter $\theta_i \in \mathbb{R}^+$ and denote by $F_\theta(.)$ its cumulative distribution function. We finally assume that the cost is separable. Basically, this cost

[13] As detailed below, the probability chosen by the interviewer differs however from the observed response rate z_{ij}.

separability assumption is a restriction that 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). Under our assumptions, the cost function satisfies:

$$C_{ij}(y) = C_i(y) = \theta_i C(y).$$

The heterogeneity θ_i is known by the interviewers. Indeed, they work for Insee in the same area and usually for a very long time (the average experience was around 6.8 years in 2001). Thus, it seems reasonable to assume that they anticipate correctly the difficulties they will face.

Insee compensates this cost and gives incentives to the interviewer through the following scheme. We denote by δ_j and w_j the bonus and basic wage for survey j. The interviewer thus receives $w_j + \delta_j$ from Insee when the interview is achieved and w_j otherwise. Hence, if he implements a probability y of conducting the survey for each household in his sample, the interviewer obtains on average a total wage of $\delta_j y + w_j$. We suppose hereafter that interviewers are risk-neutral and have a quasi-linear utility function. In this case, they solve

$$\max_y \ \delta_j y - \theta_i C(y). \tag{3.1}$$

We also impose the following mild regularity condition.

Assumption 1 $C(.)$ *is twice continuously differentiable,* $C(0) = C'(0) = 0$ *and* $C''(y) > 0$ *for all* $y \in (0, 1)$*.* $F_\theta(.)$ *is continuously differentiable with density* $f_\theta(.)$ *and support* $\mathbb{R}^+$*.*

Under this assumption, Program (3.1) admits for all θ a unique solution $y_j(\theta)$ which satisfies the first order condition $\delta_j = \theta C'(y_j(\theta))$. Moreover, differentiating this condition shows that $\theta \mapsto y_j(\theta)$ is strictly decreasing.

Finally, it is important to remember that we do not observe directly the probability $y_{ij} \equiv y_j(\theta_i)$ chosen by interviewer i for survey j but only the number of respondents r_{ij} among the n_{ij} households that he has to interview. In general, the observed response rate $z_{ij} = r_{ij}/n_{ij}$ is different from y_{ij}. In particular, the fact that n_{ij} is finite and small explains why we observe jumps in the distribution functions of z_{ij} (see Figures 1 and 2), whereas the one of y_{ij} is continuous under Assumption 1. To model the theoretical link between y_{ij}, r_{ij} and n_{ij}, we first impose Assumption 2.

Assumption 2 *n_{ij} is independent of θ_i ant its support is the set of natural integers.*

Independence between n_{ij} and θ_i was established above, as we show that the interviewer' fixed effect (which corresponds to θ_i here) is unrelated with the number of household he receives. Besides, because the interviewer contacts each of his households with probability y_{ij}, the number of respondents r_{ij} satisfies, provided that each household reacts independently from each other,

$$r_{ij}|n_{ij}, y_{ij} \sim \text{Binomial}(n_{ij}, y_{ij}).$$

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.

3.2 Insee's program

To complete the model, we have to describe how Insee chooses the contract it proposes to the interviewers. We have reasons to believe that Insee's contracts are suboptimal. This is confirmed by two facts. The first is the violation of the Informativeness Principle which states that all factors correlated with performance should be included in the contracts (Prendergast, 1999). Here, for instance, the bonus does not depend on the 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%). The second is the fact 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. Instead, we suppose that it maximizes his objective function only in the class of linear contracts.

Let $S_j(y)$ denote Insee's objective function in survey j if the response rate is y, and $y(\theta, \delta)$ be the response rate chosen by an interviewer of type θ when the bonus is δ. We have, by the optimality of the observed payment scheme among linear contracts,[14]

$$\delta_j = \arg\max_{\delta} E\left[S_j(y(\theta, \delta)) - \delta y(\theta, \delta)\right].$$

[14]We do not mention the maximization on the basic wage w_j. It is simply set such that the worse type interviewer obtains his outside utility.

Hence, δ_j satisfies[15]

$$-E\left[y(\theta,\delta_j)\right]+E\left[\frac{\partial y}{\partial \delta}(\theta,\delta_j)(S_j'(y(\theta,\delta_j))-\delta_j)\right]=0. \tag{3.2}$$

3.3 Policy analysis

Given its policy, Insee's surplus is

$$\Pi_j=E\left[S_j(y_j(\theta))-\delta_j y_j(\theta)\right].$$

This surplus is not the optimal one since Insee restricts itself to linear contracts only. To derive the optimal contract, we impose the following standard regularity condition.

Assumption 3 *$\theta\mapsto\theta+F_\theta(\theta)/f_\theta(\theta)$ is increasing.*

Under Assumption 3, the optimal contracts are defined (see, e.g., Laffont & Martimort, 2002) by the following system of equations:

$$\begin{array}{rcl} S_j'(y_j^I(\theta) & = & \left[\theta+\dfrac{F_\theta(\theta)}{f_\theta(\theta)}\right]C'(y_j^I(\theta)),\\ t_j^{I\prime}(y_j^I(\theta)) & = & \theta C'(y_j^I(\theta), \end{array}$$

where $y_j^I(\theta)$ corresponds to the response rate chosen by an interviewer of type θ, facing the optimal payment scheme $t_j^{I\prime}(.)$. Under this optimal contract, Insee's surplus satisfies

$$\Pi_j^I=E\left[S_j(y_j^I(\theta))-t_j^I(y_j^I(\theta))\right].$$

We can then compare the previous surpluses with the one Insee would obtain without asymmetric information, i.e. observing the type of each interviewer. Under complete information, Insee is able to fix the response rate interviewer by interviewer. These optimal response rates are given by

$$S_j'(y_j^C(\theta))=\theta C'(y_j^C(\theta)).$$

Moreover, Insee recovers all the rent from the interviewers. As a result, the optimal transfer function $t_j^C(.)$ is defined by

[15]Given the restriction we impose hereafter (namely, $S_j(y)=\lambda_j y$), one can show that the first order condition of the program is necessary and sufficient.

$$t_j^C(y_j^C(\theta)) = \theta C(y_j^C(\theta)),$$

and the expected surplus under complete information satisfies

$$\Pi_j^C = E\left[S_j(y_j^C(\theta)) - t_j^C(y_j^C(\theta))\right].$$

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 (3.2), where the expectations are taken on the considered populations of interviewers.

4 Inference on the model

4.1 Identification

We now turn to the empirical content of the model. We consider an ideal framework where the number of interviewers in the 2001-2002[16] and 2003 surveys (indexed by $j = 1$ and $j = 2$ respectively) 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.[17] The question is whether the marginal cost functions $C'(.)$, the distribution of types $F_\theta(.)$ and the objective functions $S_j(.)$ can be recovered from these functions and the model.

We first show that it is possible to recover the distribution of y_j from the one of (n_j, r_j). We have

$$\begin{aligned} P(r_j = n|n_j = n) &= E\left[P(r_j = n|n_j = n, y_j)\right] \\ &= E\left[y_j^n|n_j = n\right] \\ &= E\left[y_j^n\right], \end{aligned}$$

where the second equality follows from the binomial distribution assumption and the third from Assumption 2. As a result, all moments of y_j are identified since, by Assumption 2,

[16] From now on, we aggregate the 2001 and 2002 surveys, as they are identical. The number of respondents and subsample sizes on 2001-2002 are thus the sums of these two variables over the two surveys.

[17] We omit subscript i for simplicity here.

the support of n_j is the set of natural integers.[18] Because y_j is bounded, this identifies the distribution $F_{y_j}(.)$ of y_j (see, e.g., Gut, 2005).

Second, we investigate the identification of $C'(.)$ and $F_\theta(.)$. Before turning to our results, note that a normalization is necessary since for any $\alpha > 0$, we can replace $(\theta, C'(.))$ by $(\alpha\theta, C'(.)/\alpha)$ and leave the model unchanged. Indeed, the economic model is not completely specified. All models with the same total cost will be equivalent and one can always increase θ and decrease $C'(.)$ accordingly without modifying the economic model. Hence, for a given $\theta_0 > 0$, we can choose any y_0 in $(0, 1)$ such that $\theta_1(y_0) = \theta_0$.[19]

We also introduce two types of transforms that are at the basis of our identification method in the presence of an exogenous change. First, let us consider an interviewer of type $\widetilde{\theta}$. His theoretical response rate is $y_1(\widetilde{\theta})$ in survey 1 and $y_2(\widetilde{\theta})$ in survey 2. Because these theoretical response rates are decreasing with θ, their rank in the distributions $F_{y_1}(.)$ and $F_{y_2}(.)$ are identical:

$$F_{y_1}(y_1(\widetilde{\theta})) = \mathbb{P}(y_1(\theta) \leq y_1(\widetilde{\theta})) = \mathbb{P}(\theta \geq \widetilde{\theta}) = F_{y_2}(y_2(\widetilde{\theta})). \tag{4.1}$$

Introducing the horizontal transform $H_{jk}(.)$ defined by $H_{jk}(y) = F_{y_k}^{-1}\left[F_{y_j}(y)\right]$, we get

$$y_k(\widetilde{\theta}) = H_{jk}(y_j(\widetilde{\theta})). \tag{4.2}$$

As the distribution functions $F_{y_j}(.)$ are identified, $H_{jk}(.)$ also is, and the knowledge of $y_j(\widetilde{\theta})$ implies the knowledge of $y_k(\widetilde{\theta})$. From an economic perspective, this equality simply states that it is possible to recover the theoretical response rate of an interviewer of type $\widetilde{\theta}$ in survey k if we know which production he chooses in survey j. To do so, even if his type $\widetilde{\theta}$ is unobserved, it is sufficient to pick the quantile of F_{y_k} corresponding to $F_{y_j}(y_j(\widetilde{\theta}))$.

We also rely on the agent's program by using his first order condition. Letting $\theta_j(.)$ denote the inverse function of $y_j(.)$, we have

$$\delta_j = \theta_j(y)C'(y). \tag{4.3}$$

Hence, defining the vertical transform $V_{jk}(.)$ by $V_{jk}(\theta) = \frac{\delta_k}{\delta_j}\theta$, we obtain, for all $y \in (0, 1)$,

$$\theta_k(y) = V_{jk}(\theta_j(y)). \tag{4.4}$$

[18]In the data, $\max_i n_{i1} = 50$ and $\max_i n_{i2} = 53$, which ensures the identification of more than 50 moments of the distribution.

[19]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 (4.3) below), which in turn applies to $\theta_2(y_0)$.

$V_{jk}(.)$ is identified, so that the knowledge of $\theta_j(y)$ implies the knowledge of $\theta_k(y)$. Contrary to the horizontal transform which links different response rates that similar interviewers choose in both surveys, the vertical transform links different types of interviewer who chooses the same level of response rate in both surveys. Knowing the type of an interviewer with an optimal response rate of y in survey k, it is possible to recover the type of the interviewer that chooses the same level y in survey j.

Figure 4 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 $(y_0, \theta_1(y_0))$, we can identify $(y_1, \theta_1(y_1))$ where $y_1 = H_{12}(y_0)$ and $\theta_1(y_1) = V_{21}(\theta_1(y_0))$. By induction, we identify all the black points in Figure 4.

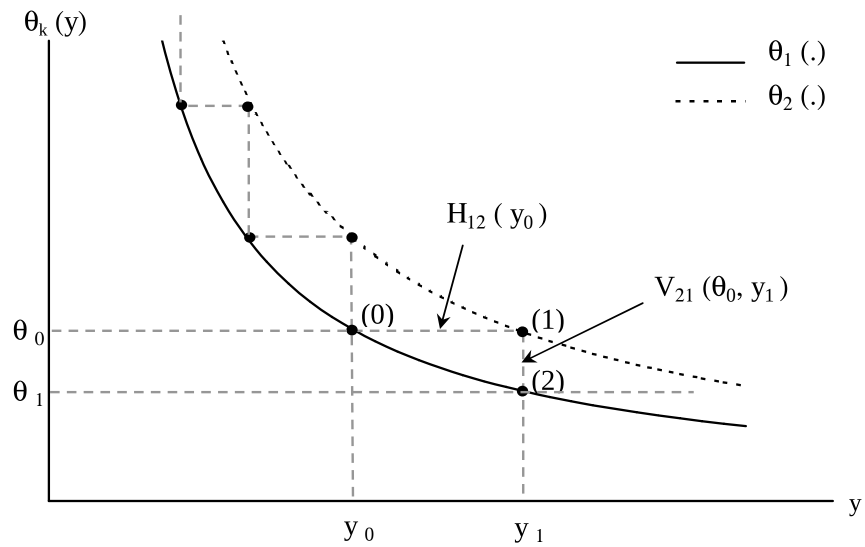

Figure 4: The horizontal and vertical transforms.

Formally, let $H_{12}^n(y) = H_{12} \circ ... \circ H_{12}(y)$ if $n > 0$, y if $n = 0$ and $H_{21} \circ ... \circ H_{21}(y)$ if $n < 0$. We identify $\theta_1(.)$ on the increasing sequence $(y_n)_{n \in \mathbb{Z}}$ defined by $y_n = H_{12}^n(y_0)$.[20] Elsewhere, $\theta_1(.)$ can be bounded, using the property that it is a decreasing function. Finally, using Equation (4.1) and the first order condition (4.3), we obtain bounds on $F_\theta(.)$ and $C'(.)$. Theorem 4.1 makes these bounds explicit and show that they are sharp.[21]

[20]That $(y_n)_{n \in \mathbb{Z}}$ is increasing follows from the fact that $H_{12}(y) > y$ for all $y \in (0, 1)$ (since $y_1(.) < y_2(.)$).

[21]Note that the bounds are pointwise sharp but not functionally sharp.

Theorem 4.1 *Suppose that Assumptions 1 and 2 hold. Then, for all* $y \in (0,1)$ *and all* $\theta > 0$, $C'(y) \in [\underline{C'}(y), \overline{C'}(y)]$ *and* $F_\theta(\theta) \in [\underline{F}_\theta(\theta), \overline{F}_\theta(\theta)]$, *with*

$$\underline{C'}(y) = \frac{\delta_1}{\inf_{n\in\mathbb{Z}:\, y_n \le y} \theta_1(y_n)}, \qquad \overline{C'}(y) = \frac{\delta_1}{\sup_{n\in\mathbb{Z}:\, y_n \ge y} \theta_1(y_n)}, \tag{4.5}$$

$$\underline{F_\theta}(\theta) = 1 - F_{y_1}\left(\inf_{n\in\mathbb{Z}:\, \theta_1(y_n)\le\theta} y_n\right), \qquad \overline{F_\theta}(\theta) = 1 - F_{y_1}\left(\sup_{n\in\mathbb{Z}:\, \theta_1(y_n)\ge\theta} y_n\right). \tag{4.6}$$

These bounds are identified and sharp. Finally, for all $n \in \mathbb{Z}$, $\underline{C'}(y_n) = \overline{C'}(y_n)$ *and* $\underline{F}_\theta(\theta_1(y_n)) = \overline{F}_\theta(\theta_1(y_n)))$. *Thus,* $C'(.)$ *and* $F_\theta(.)$ *are point identified respectively on the sequences* $(y_n)_{n\in\mathbb{Z}}$ *and* $(\theta_1(y_n))_{n\in\mathbb{Z}}$.

Theorem 4.1 provides the best nonparametric bounds on the agents' cost function and distribution of heterogeneity. Our identification result strongly relies on the use of an exogenous change. In particular, without variations in the contracts (i.e., when we observe data from only one menu of contract or if the change is endogenous), one can prove that the model is not identified.[22] Any increasing marginal cost function $C'(.)$ (and similarly, any distribution function $F_\theta(.)$) can be rationalized by the data. Similar results have also 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. More generally, exogenous changes are necessary, and sometimes sufficient, to completely identify any basic adverse selection model (see D'Haultfœuille & Février, 2010). This framework includes regulation, nonlinear pricing / price discrimination models, financial contracts or simple insurance settings. All these models share a common underlying structure (see Laffont & Martimort, 2002 for a survey) and the method proposed here applies similarly for all these applications.

Our result also implies that standard parametric models on $C'(.)$ and $F_\theta(.)$ are identified with an exogenous change. For instance, the parameters of a lognormal, weibull or gamma distribution are identified thanks to the knowledge of $F_\theta(.)$ on the sequence $(\theta_1(y_n))_{n\in\mathbb{Z}}$. Actually, because we retrieve an infinite sequence of points on $C'(.)$ and $F_\theta(.)$, such standard parametric models are overidentified. The sequences $(C'(y_n))_{n\in\mathbb{Z}}$ and $(F_\theta(\theta_1(y_n)))_{n\in\mathbb{Z}}$ may thus serve as a guidance for choosing appropriate parametric restrictions, as will be the case in Section 5.

Investigating the identification of $S_j(.)$, it is clear that Equation (3.2) does not allow to identify nonparametrically the whole function of interest. Even if we supposed $F_\theta(.)$ and

[22] The formal proof of non-identification is established in Appendix A.

$C'(.)$ (and consequently $\frac{\partial y}{\partial \delta}(.,.)$) to be known, we would recover only one parameter, namely the mean of its derivative. We thus restrict ourselves to the class of linear functions of the form $S_j(y) = \lambda_j y$. Under this specification, λ_j represents the "price" of the information contained in a household's answers, i.e. the social value of an interview in survey j. In our framework, $\lambda_1 < \lambda_2$ as Insee values more the 2003 answers than those of 2001 or 2002. For given $F_\theta(.)$ and $C'(.)$, λ_j is just identified and satisfies

$$\lambda_2 = \delta_2 + \frac{E(y_2)}{E\left(\frac{\partial y}{\partial \delta}(\theta, \delta_2)\right)}.$$

4.2 Estimation

We now turn to the estimation of $C'(.)$, $F_\theta(.)$ and $S_j(.)$. We let $\mathcal{S}_j$ denote the sample of interviewers participating to survey j, N_j the corresponding sample size and $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. For ease of notation, we let $r_{ij} = n_{ij} = 0$ if an interviewer i does not participate to survey j. We study the behavior of our estimators when $N = \min(N_1, N_2) \to \infty$ and under the following standard assumption of independent sampling.

Assumption 4 *(independent sampling)* $(\theta_i, r_{i1}, n_{i1}, r_{i2}, n_{i2})_{i \in \mathcal{S}}$ *are i.i.d.*

We address first the nonparametric estimation of bounds on $C'(.)$ and $F_\theta(.)$. We consider then the parametric estimation of both functions, as well as the parametric estimation of $S_j(.)$.

4.2.1 Nonparametric estimation

Our nonparametric estimation method follows closely the identification strategy and may be decomposed into two steps. We first estimate the distributions $F_{y_1}(.)$ and $F_{y_2}(.)$ of the unobserved probabilities $y_1(\theta)$ and $y_2(\theta)$. We then estimate bounds on the primitive functions $C'(.)$ and $F_\theta(.)$.

For the first step, we use a sieve maximum likelihood estimator (see, e.g., Chen, 2006, for a survey on sieve estimation). More precisely, we choose to approximate the densities[23] $f_{y_1}(.)$ and $f_{y_2}(.)$ by functions of the sieve space

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

[23] Assumption 1 and Equation (4.1) ensure that the densities of y_1 and y_2 do exist.

where 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 which tends to infinity. We thus approximate the density $f_{y_j}(.)$ 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.[24] Indeed, 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 and n_j,

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

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

We now turn to the estimation of $C'(.)$ and $F_\theta(.)$. First, letting $\widehat{H}_{jk}(x) = \widehat{F}_{y_k}^{-1} \circ \widehat{F}_{y_j}(x)$, we estimate y_n by $\widehat{y}_n = \widehat{H}_{12}^n(y_0)$. Note that because $V_{21}(\theta) = \delta_1/\delta_2 \times \theta$, $\theta_n = \theta_1(y_n) = (\delta_1/\delta_2)^n \theta_0$ and does not need to be estimated. Then, relying on (4.5) and (4.6), the bounds on $C'(.)$ and $F_\theta(.)$ are estimated by

$$\begin{aligned}
\widehat{\underline{C'}}(y) &= \frac{\delta_1}{\inf_{n\in\mathbb{Z}:\,\widehat{y}_n \le y} \theta_n}, \qquad \widehat{\overline{C'}}(y) = \frac{\delta_1}{\sup_{n\in\mathbb{Z}:\,\widehat{y}_n \ge y} \theta_n}, \\
\widehat{\underline{F_\theta}}(\theta) &= 1 - \widehat{F}_{y_1}\left(\inf_{n\in\mathbb{Z}:\,\theta_n \le \theta} \widehat{y}_n\right), \qquad \widehat{\overline{F_\theta}}(\theta) = 1 - \widehat{F}_{y_1}\left(\sup_{n\in\mathbb{Z}:\,\theta_n \ge \theta} \widehat{y}_n\right).
\end{aligned}$$

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.

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

Assumption 5 $\lim_{\theta\to\infty}\theta^2 f(\theta) = 0$ *and* $\lim_{y\to 1}\frac{C''(y)}{C'(y)^2}$ *exists and is finite. For all* $u > 0$ *and* $j \in \{1,2\}$, $E(u^{n_j}) < \infty$.

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 4.2 shows that these conditions are sufficient for the consistency of the estimated bounds.

Theorem 4.2 *Suppose that Assumptions 1, 2, 4 and 5 hold,* $K_N \to \infty$ *and* $K_N^2 \ln K_N/N \to 0$. *Then* $\widehat{\underline{F}}_\theta(\theta)$ *and* $\widehat{\overline{F}}_\theta(\theta)$ *are consistent for all* $\theta > 0$. $\widehat{\underline{C'}}(y)$ *and* $\widehat{\overline{C'}}(y)$ *are consistent on every* $y \notin \{y_n, n \in \mathbb{Z}\backslash\{0\}\}$. *Moreover, for all* $n \in \mathbb{Z}$,

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

Theorem 4.2 has three parts. The first establishes consistency of the bounds of $F_\theta(.)$ on its whole support. The second shows the convergence of $\underline{C}'(.)$ and $\overline{C}'(.)$ outside the sequence $(y_n)_{n\in\mathbb{Z}}$. 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 $\left(\widehat{y}_n, \widehat{\overline{C'}}(\widehat{y}_n)\right)$. As a consequence $C'(.)$ and $F_\theta(.)$ are well estimated on the sequences where they are point identified, while sharp bounds are consistently recovered anywhere else.

4.2.2 Parametric estimation

The nonparametric estimation is not sufficient to conduct the policy analysis detailed in Subsection 3.3. Both $F_\theta(.)$ and $C'(.)$ have to be known on their full support. Hence, we also consider parametric restrictions on $F_\theta(.)$ and $C'(.)$. We write these functions as $F_\theta(.|\eta)$ and $C'(.|\eta)$, for an unknown finite dimensional parameter η. In this case,

$$y_j(\theta|\eta) = C'^{-1}\left(\frac{\delta_j}{\theta}|\eta\right). \tag{4.7}$$

Hence, the probability of observing (r_1, r_2) conditional on (n_1, n_2) satisfies

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

and η can then be estimated by maximum likelihood on $\mathcal{S}$.

Finally, concerning the estimation of Insee's objective functions, we estimate λ_j by

$$\widehat{\lambda}_j = \delta_j + \frac{\int y_j(\theta|\widehat{\eta}) f_\theta(\theta|\widehat{\eta}) d\theta}{\int \frac{\partial y}{\partial \delta}(\theta, \delta_j|\widehat{\eta}) f_\theta(\theta|\widehat{\eta}) d\theta},$$

where, using Equation (4.7),

$$\frac{\partial y}{\partial \delta}(\theta, \delta_j|\widehat{\eta}) = \frac{1}{\theta C''\left(C'^{-1}\left(\frac{\delta_j}{\theta}|\widehat{\eta}\right)\right)}.$$

Similarly, for the policy analysis, all surpluses defined in Subsection 3.3 are estimated using the parametric restriction we consider and the estimated parameter $\widehat{\eta}$.[25]

5 Results

5.1 Estimation of $C'(.)$, $F_\theta(.)$ and λ_j.

We first estimate nonparametrically the sharp bounds on $F_\theta(.)$ and $C'(.)$. For that purpose, we estimate in a first step $F_{y_1}(.)$ and $F_{y_2}(.)$ 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 $K_N = 4$. The corresponding estimates are displayed in Figure 5. As predicted by the theory, the distribution function of y on the 2003 survey dominates stochastically the one of 2001-2002 on most part of $(0, 1)$.

[25] The estimators defined here can be obtained either by using closed-form formulas for the integrals or by simulations, depending on the parametric choice of $C'(.)$ and $F_\theta(.)$.

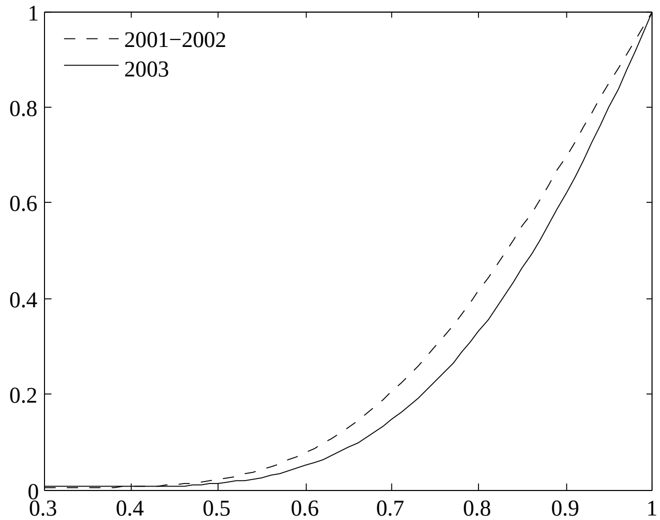

Figure 5: Sieve MLE Estimates of $F_{y_1}(.)$ and $F_{y_2}(.)$.

In the second step, we estimate the bounds on $F_\theta(.)$ and $C'(.)$. We choose a starting value y_0 close to the median of $\widehat{F}_{y_1}(.)$, namely $y_0 = 0.8$, in order to get more precise estimates for central values of $F_\theta(.)$ and $C'(.)$.[26] For that y_0, we impose the normalization $\theta_1(y_0) = 1$. Figure 6 displays the estimates of the bounds on $F_\theta(.)$ 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. This may explain the small average effect of incentives that we have found compared to the previous results of the literature. Finally, the width of the confidence intervals on the bounds of $F_\theta(.)$ (resp. $C'(.)$) increases with $|\theta - 1|$ (resp. $|y - 0.8|$), reflecting the fact that, as expected, the estimation error increases with $|n|$.

[26]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.

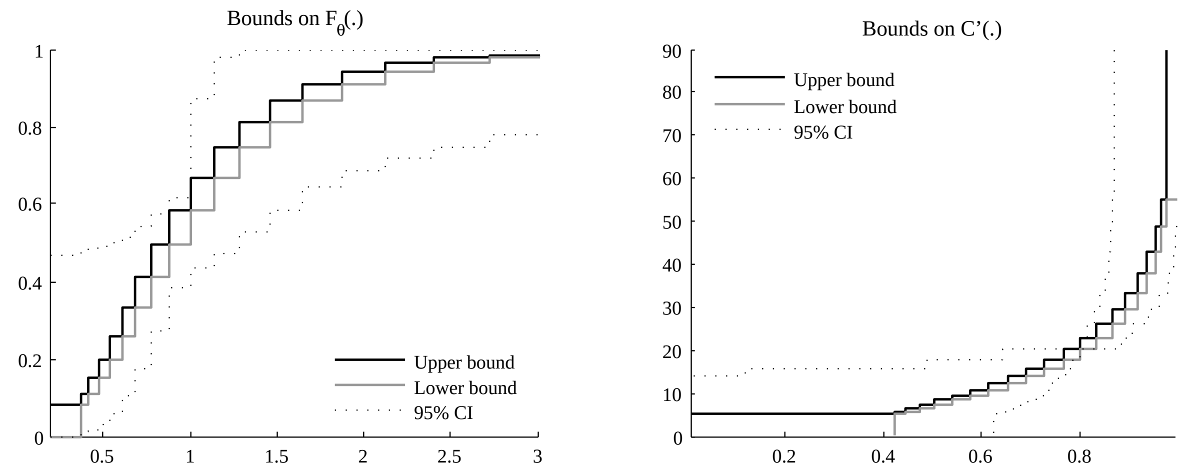

Figure 6: Estimated bounds on $F_\theta(.)$ and $C'(.)$.

The nonparametric approach is appealing as it reveals what can be identified when imposing only the exogeneity of the change. Yet, it does not allow us to compute parameters of interest such as surpluses under optimal contracts or symmetric information. Thus, we also consider a parametric estimation of $F_\theta(.)$ and $C'(.)$. For this purpose, we use the nonparametric estimates $(C'(y_n), F_\theta(\theta_n))_{n\in\mathbb{Z}}$ to investigate which parametric family fits best. We compare three standard family of distributions on $\mathbb{R}^+$ for $F_\theta(.)$, namely the Fréchet, for which $F_\theta(\theta) = \exp(-a\theta^{-b})$ $(a, b > 0)$, the lognormal, for which $F_\theta(\theta) = \Phi((\ln\theta - a)/b)$ (where $\Phi(.)$ denotes the cumulative distribution function of a standard normal variable and $b > 0$) and the Weibull, for which $F_\theta(\theta) = 1 - \exp(-a\theta^b)$ $(a, b > 0)$. 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 $\ln(-\ln F_\theta(\theta_n))$, $\Phi^{-1}(F_\theta(\theta_n))$ and $\ln(-\ln(1 - F_\theta(\theta_n)))$ against $\ln\theta_n$. Points should be aligned if the parametric family is the true one. Similarly, we consider families of marginal cost functions tending to 0 at 0 and to ∞ at 1, but which differ in their behavior at infinity. More precisely, 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 $\ln C'(y_n)$ against $\ln\phi(y/(1-y))$ in the three cases.

Figures 7 and 8 display the three corresponding plots. They indicate that the lognormal distribution and $\phi(x) = \ln(1+x)$ have the best fits. 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 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 $F_\theta(.)$. 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. Maximum likelihood estimates of the parameters for this specification are displayed in Table 8. We obtain in a second step $\widehat{\lambda}_1 = 84.4$ and $\widehat{\lambda}_2 = 104.1$, the higher value of $\widehat{\lambda}_2$ reflecting the higher importance for Insee of the 2003 survey.

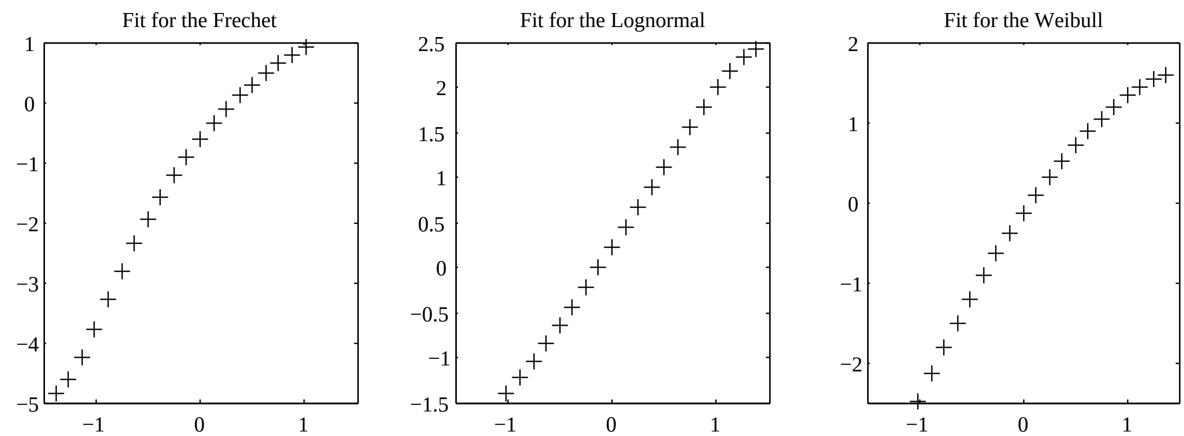

Figure 7: Choice of the parametric family for $F_\theta(.)$.

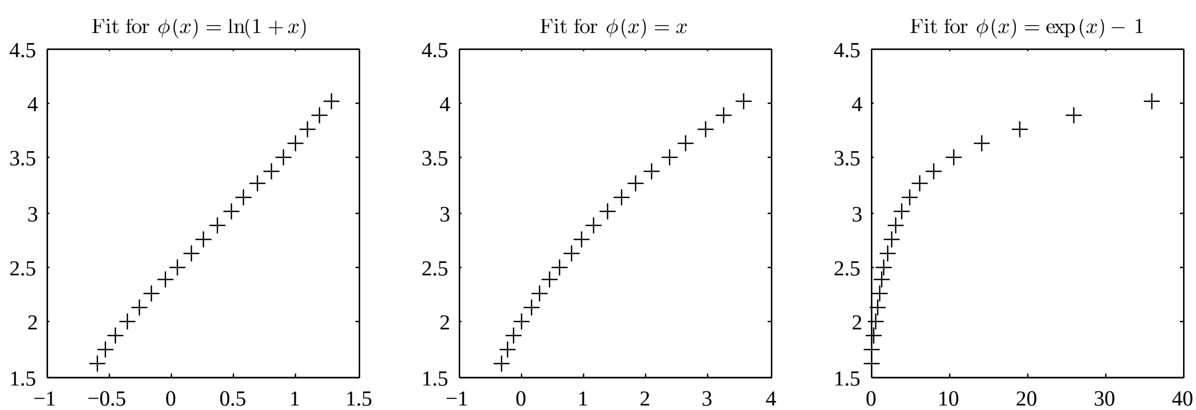

Figure 8: Choice of the parametric family for $C'(.)$.

Parameter	Estimate
a	-0.03 (0.04)
b	0.45 (0.13)
α	11.57 (1.64)
β	1.17 (0.34)
λ_1	84.4 (29.3)
λ_2	104.1 (22.5)

Table 8: Maximum likelihood estimates of the parameters of $C'(y) = \alpha\,[\ln(1 + y/(1-y))]^{\beta}$ and $F_\theta(\theta) = \Phi((\ln\theta - a)/b)$.

5.2 The cost of using inefficient contracts

We now turn to the results on surpluses. We focus on the 2003 survey, the results being very similar for 2001-2002. Table 9 summarizes our results. We find that the surplus loss associated with the use of linear contracts is around 16% (62.3 versus 74.4) 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. A recent empirical paper of Miravete (2007), which reports a loss of only 3%, also supports this claim. These empirical results are in line 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. (2008) 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.

We find moderate cost of incomplete information, the optimal surplus under asymmetric

Environment	Pay method	E[surplus]	Relative	E[response rate]
Full information	Optimal contract	93.9 (19.8)	1.00 (0)	99% (0.01)
Incomplete information	Optimal contract	74.4 (19.7)	0.79 (0.03)	93% (0.01)
Incomplete information	Linear contract	62.3 (18.3)	0.66 (0.04)	83% (0.01)

Table 9: Surplus and response rates under alternative compensation schemes.

information being 79% of the optimal one under full information. This loss of 21% 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 (62%) is due to incomplete information whereas 38% is associated with the simple tarification.

			Ratio of surplus			E[response rate]		
F_θ	$\phi(x)$	λ_2	Optimal	Linear	Lin/Opt	Full info.	Optimal	Linear
	$\ln(1+x)$	108.0	0.83	0.71	0.85	0.99	0.91	0.82
Fréchet	x	103.0	0.80	0.70	0.88	0.98	0.90	0.82
	$\exp(x)-1$	115.6	0.77	0.72	0.94	0.94	0.86	0.82
	$\ln(1+x)$	104.1	0.79	0.66	0.84	0.99	0.93	0.83
Lognormal	x	100.3	0.76	0.66	0.87	0.99	0.92	0.83
	$\exp(x)-1$	112.1	0.75	0.70	0.94	0.94	0.88	0.82
	$\ln(1+x)$	101.7	0.68	0.64	0.95	1.00	0.83	0.82
Weibull	x	98.7	0.65	0.65	0.99	0.99	0.88	0.83
	$\exp(x)-1$	110.5	0.71	0.70	0.98	0.94	0.88	0.83

Table 10: Robustness of the results with respect to the parametric families.

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 distinguished, for instance, between interviewers in large urban areas and other interviewers.[27] Instead of offering the same linear contract with a bonus of 22.9 euros to all interviewers, it would propose in this case a bonus of 24.3 euros in large urban areas and a bonus of 22.1 euros elsewhere. The expected surplus would remain however nearly constant (62.5 instead of 62.3). The cost

[27] There are 167 interviewers in large urbain areas defined as towns larger than 100,000 inhabitants and 286 interviewer in other areas.

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.

To conclude, we assess the importance of using the exogenous change and our nonparametric analysis in the choice of the parametric specification. To do so, we compute the surplus and average response rates under nine parametric families. As previously, we use the Fréchet, lognormal and Weibull distributions for the types and $\phi(x) = \ln(1 + x)$, x or $\exp(x) - 1$ for the marginal cost. The results, displayed in Table 10, appear to be stable for the Fréchet or lognormal combined with $\phi(x) = \ln(1+x)$ or x. For these specifications, the cost of asymmetric information is moderate, around 20% (between 17 and 24%), while the cost of using simple contracts is small, between 12 and 16%. The pattern is however quite different when we choose either the Weibull distribution or $\phi(x) = \exp(x) - 1$. For instance, with the Weibull distribution and $\phi(x) = x$, the cost of asymmetric information increases to 35%, whereas there is almost no loss of using linear contracts (1% only). These results are in line with the plots displayed in Figures 7 and 8, where the Fréchet and the lognormal distributions on the one hand, and $\phi(x) = \ln(1+x)$ and x on the other hand display correct fits, while the Weibull distributions and $\phi(x) = \exp(x) - 1$ seem less appropriate. This analysis emphasizes the importance of avoiding any parametric misspecification, which can be achieved only through exogenous changes and our nonparametric method.

6 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

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. (2008), 'Nearly optimal pricing for multiproduct firms'. NBER Working Paper, 13916.

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. (2009), Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. Mimeo.

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.

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.

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

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.

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

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.

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.

Misra, S. & Nair, H. (2008), A structural model of sales-force compensation dynamics: Estimation and field implementation. Mimeo.

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

Paarsch, L. & Shearer, B. (1999), 'The response of worker effort to piece rates: Evidence from the british columbia tree-planting industry', *Journal of Human Resources* **34**, 643–667.

Paarsch, L. & 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. (2011), 'Nonparametric identification of a contract model with adverse selection and moral hazard', *Econometrica, Forthcoming* .

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.

Shearer, B. (2004), ‘Piece rates, fixed wages and incentives: Evidence from a field experiment’, *Review of Economic Studies* **71**, 513–534.

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

Steenburgh, T. J. (2008), ‘Effort or timing: The effect of lump-sum bonuses’, *Quantitative Marketing and Economics* **6**, 235–256.

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.

Appendix A: proofs

Proof of Theorem 4.1

It follows from the discussion before Theorem 4.1 that $\theta_1(.)$ is point identified on $(y_n)_{n\in\mathbb{Z}}$. For other y, we get, by monotonicity of $\theta_1(.)$,

$$\sup_{n\in\mathbb{Z}:\, y_n\geq y} \theta_1(y_n) \leq \theta_1(y) \leq \inf_{n\in\mathbb{Z}:\, y_n\leq y} \theta_1(y_n).$$

Similarly,

$$\sup_{n\in\mathbb{Z}:\, \theta_1(y_n)\geq\theta} y_n \leq y_1(\theta) \leq \inf_{n\in\mathbb{Z}:\, \theta_1(y_n)\leq\theta} y_n.$$

By Equations (4.1) and (4.3), Inequalities (4.5) and (4.6) 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 $y\in(0,1)$ ($y\notin(y_n)_{n\in\mathbb{Z}}$) and $\theta>0$ ($\theta\notin(\theta_1(y_n))_{n\in\mathbb{Z}}$), the bounds on $C'(y)$ and $F_\theta(\theta)$ are sharp. We focus, for a given y, on $\overline{C'}(y)$ as the proof is similar for $\underline{C}'(y)$, $\overline{F}_\theta(\theta)$ and $\underline{F}_\theta(\theta)$. 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 $\widetilde{\theta}_1(y_n)=\theta_1(y_n)$ for all $n\in\mathbb{Z}$ 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

$$\lim_{y\to y_{k+1}^-}\widetilde{\theta}'_1(y)=\frac{\delta_1\widetilde{\theta}'_1(y_k)}{\delta_2 H'_{12}(y_k)}. \tag{6.1}$$

Such a function always exists. We then extend it on $(0,1)$ through the vertical and horizontal transforms. For instance, $\widetilde{\theta}_1(.)$ is defined on $[y_{k+1},y_{k+2}[=[H_{12}(y_k),H_{12}(y_{k+1})[$ by

$$\widetilde{\theta}_1(y)=\frac{\delta_1}{\delta_2}\widetilde{\theta}_1(H_{12}^{-1}(y)).$$

Moreover, because $H_{12}(.)$ is continuously differentiable, $\widetilde{\theta}_1(.)$ admits a right derivative at y_{k+1} given by

$$\lim_{y \to y_{k+1}^+} \widetilde{\theta}_1'(y) = \frac{\delta_1 \widetilde{\theta}_1'(y_k)}{\delta_2 H_{12}'(y_k)},$$

and Equation (6.1) ensures that $\widetilde{\theta}_1(.)$ is differentiable at y_{k+1}. By induction, using either $H_{12}(.)$ or $H_{21}(.)$, 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 $H_{12}(.)$ or $H_{21}(.)$ are increasing.

We now consider the function $\widetilde{C}'(.) = \delta_1/\widetilde{\theta}_1(.)$. By construction, the first order condition $\widetilde{\theta}_1(.)\widetilde{C}'(.) = \delta_1$ 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,

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

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.[28] ■

Non-identification with one menu of contracts

Let us consider a strictly increasing and differentiable function $\widetilde{C}'(.)$, different from the true one C'. Define then $\widetilde{\theta}(.)$ by

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

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

$$\widetilde{F}_\theta(\theta) = F_{y_1}(\widetilde{y}(\theta)).$$

By construction $\widetilde{C}'(.)$ and $\widetilde{F}_\theta(.)$ are consistent with the first and second order conditions and the identified distribution $F_{y_1}(.)$. As a result, $C'(.)$ and $F_\theta(.)$ are not identified. ■

[28] Theoretically, we should also check that, in association with the considered $\widetilde{C}'(.)$, there exists a function $F_\theta(.)$ satisfying all the constraints. It is easy to see that $F_{\widetilde{\theta}}(.) = 1 - F_{y_1}(\widetilde{y}_1(.))$, where $\widetilde{y}_1(.)$ is the inverse of $\widetilde{\theta}(.)$, works.

Proof of Theorem 4.2

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

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

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 $j \in \{1,2\}$,

$$\left\| \widehat{f}_{y_j} - f_{y_j} \right\| \xrightarrow{\mathbb{P}} 0. \tag{6.2}$$

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

$$\theta_j'(y) = -\frac{\theta_j(y)C''(y)}{C'(y)} = -\frac{\delta_j C''(y)}{C'^2(y)}. \tag{6.3}$$

Thus, by Assumption 5, $\lim_{y\to 1} f_{y_j}(y)$ exists and is finite. The same holds at 0. Thus, we can extend $f_{y_j}(.)$ 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 $Q_j(f) = E(l(f, r_j, n_j))$ denote the expectation of $l(f, r, n)$ with respect to (r_j, n_j) and

$$Q_{N_j,j}(f) = \frac{1}{N_j} \sum_{i=1}^{N_j} l(f, r_{ij}, n_{ij}).$$

By definition of $\widehat{f}_{y_j}$, $\widehat{f}_{y_j} = \arg\max_{f\in\mathcal{F}_N} Q_{N_j,j}(f)$ is a sieve M-estimator. We use Theorem 3.1 of Chen (2006) and its associated Remark 3.2 to prove (6.2). To this end, we check the following conditions:

a. Q_j is uniquely maximized at f_{y_j} and $Q_j(f_{y_j}) > -\infty$.
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_j is continuous for $\|.\|$.
d. $\mathcal{F}_N$ is compact.
e. $E\left[\sup_{f\in\mathcal{F}_N} |l(f, r_j, n_j)|\right] < \infty$.
f. There exists $U(.,.)$ such that $E(U(r_j, n_j)) < \infty$ 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 δ-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 $g \in \mathcal{F}$,

$$\begin{aligned}
E\left[\frac{\exp l(g, r_j, n_j)}{\exp l(f_{y_j}, r_j, n_j)}\middle| n_j = n\right] &= \sum_{k=0}^{n} P(r_j = k|n) \frac{\begin{pmatrix} k \\ n \end{pmatrix} \int_0^1 y^k (1-y)^{n-k} g(y) dy}{P(r_j = k|n)} \\
&= \int_0^1 \left(\sum_{k=0}^{n} \begin{pmatrix} k \\ n \end{pmatrix} y^k (1-y)^{n-k} \right) g(y) dy \\
&= \int_0^1 g(y) dy \\
&= 1.
\end{aligned}$$

Thus,

$$E\left[\frac{\exp l(g, r_j, n_j)}{\exp l(f_{y_j}, r_j, n_j)}\right] = 1.$$

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

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

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

$$\begin{aligned}
\int_0^1 y^r (1-y)^{n-r} f_{y_j}(y) dy &\geq \int_{u_1}^{1-u_1} y^r (1-y)^{n-r} f_{y_j}(y) dy \\
&\geq u_1^n \int_{u_1}^{1-u_1} \left(\frac{y}{u_1}\right)^r \left(\frac{1-y}{u_1}\right)^{n-r} f_{y_j}(y) dy \\
&\geq u_1^n \int_{u_1}^{1-u_1} f_{y_j}(y) dy \\
&\geq \frac{u_1^n}{2}. \qquad (6.4)
\end{aligned}$$

As a result, $Q_j(f_{y_j}) \geq E(n) \ln u_1 - \ln 2$. By Assumption 5, $E(n) < \infty$, so that $Q_j(f_{y_j}) > -\infty$.

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,

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

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 ε.

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

$$\left|\int_0^1 y^r(1-y)^{n-r}f(y)dy - \int_0^1 y^r(1-y)^{n-r}g(y)dy\right| \leq \|f-g\| \leq \varepsilon. \tag{6.5}$$

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 (6.4), we get

$$\int_0^1 y^r(1-y)^{n-r}f(y)dy \wedge \int_0^1 y^r(1-y)^{n-r}g(y)dy \geq \frac{u_2^n}{2}. \tag{6.6}$$

Besides, for all $a, b > 0$, $|\ln b - \ln a| \leq |b-a|/a \wedge b$. Hence, using (6.5) and (6.6), we get, for all $n \in \mathbb{N}$ and $r \in \{0, ..., n\}$,

$$\begin{aligned} |l(f,r,n) - l(g,r,n)| &= \left|\ln\left(\int_0^1 y^r(1-y)^{n-r}f(y)dy\right) - \ln\left(\int_0^1 y^r(1-y)^{n-r}g(y)dy\right)\right| \\ &\leq \frac{\left|\int_0^1 y^r(1-y)^{n-r}f(y)dy - \int_0^1 y^r(1-y)^{n-r}g(y)dy\right|}{\left(\int_0^1 y^r(1-y)^{n-r}f(y)dy\right) \wedge \left(\int_0^1 y^r(1-y)^{n-r}g(y)dy\right)} \\ &\leq \frac{2\varepsilon}{u_2^n}. \end{aligned} \tag{6.7}$$

As a result,

$$|Q_j(f) - Q_j(g)| \leq E\left|l(f,r_j,n_j) - l(g,r_j,n_j)\right| \leq 2\varepsilon E\left(\frac{1}{u_2^n}\right).$$

The expectation is finite by Assumption 5. Hence, $Q_j(.)$ 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

$$m(n,r) = \inf_{g \in \mathcal{F}_N} \int_0^1 y^r(1-y)^{n-r}g(y)dy \geq \frac{u_3^n}{2}. \tag{6.8}$$

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,

$$\begin{aligned} E\left[\sup_{f \in \mathcal{F}_N} |l(f,r_j,n_j)|\right] &\leq E\left[|\ln m(n_j,r_j)|\right] \\ &\leq E\left[|\ln 2| + n|\ln u_3|\right], \end{aligned}$$

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

f. Using (6.8) and a similar argument as in (6.7), 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 5, 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),

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

$\mathcal{F}_N$ 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

$$\ln N_b(\delta,\mathcal{G}_N,\|.\|)\leq C_0K_N^2\ln K_N.$$

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 (6.2).

2. Uniform consistency of H_{kj}.

We now establish that for all $(j,k)\in\{1,2\}^2$ and all $0<\underline{x}<\overline{x}<1$,

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

We first prove that for any compact K strictly included in $(0,1)$,

$$\sup_{y\in K}|\widehat{F}_{y_k}^{-1}(y)-F_{y_k}^{-1}(y)|\xrightarrow{\mathbb{P}}0 \tag{6.9}$$

By Assumption 1, $\theta_k'(y)<0$ and $f_\theta(\theta_k(y))>0$ for all $y\in(0,1)$. Hence, by continuity of f_θ and $\theta'(.)$, for all compact K included in $(0,1)$,

$$\min_{y\in K}f_{y_k}(y)=\min_{y\in K}[-f_\theta(\theta_k(y))\theta_k'(y)]>0. \tag{6.10}$$

If $\varepsilon>0$ is such that $E=\{x\in\mathbb{R}:\exists y\in F_{y_1}^{-1}(K):|x-y|\leq\varepsilon\}$ is a subset of $(0,1)$, (6.10) implies that $C_1=\min_{y\in E}f_{y_k}(y)>0$. Moreover, by the mean value theorem, for all $y\in K$,

$$F_{y_1}(F_{y_1}^{-1}(y)-\varepsilon)+C_1\varepsilon\leq F_{y_1}(F_{y_1}^{-1}(y))\leq F_{y_1}(F_{y_1}^{-1}(y)+\varepsilon)-C_1\varepsilon.$$

Consequently,

$$
\begin{aligned}
& \mathbb{P}\left(\sup_{y\in K}|\widehat{F}_{y_1}^{-1}(y)-F_{y_1}^{-1}(y)|>\varepsilon\right) \\
= & \ \mathbb{P}\left(\exists y\in K: \widehat{F}_{y_1}^{-1}(y)>F_{y_1}^{-1}(y)+\varepsilon \text{ or } \widehat{F}_{y_1}^{-1}(y)<F_{y_1}^{-1}(y)-\varepsilon\right) \\
= & \ \mathbb{P}\left(\exists y\in K: \widehat{F}_{y_1}(\widehat{F}_{y_1}^{-1}(y))=F_{y_1}(F_{y_1}^{-1}(y))>\widehat{F}_{y_1}(F_{y_1}^{-1}(y)+\varepsilon) \text{ or } F_{y_1}(F_{y_1}^{-1}(y))<\widehat{F}_{y_1}(F_{y_1}^{-1}(y)-\varepsilon)\right) \\
\leq & \ \mathbb{P}\left(\exists y\in K: F_{y_1}(F_{y_1}^{-1}(y)+\varepsilon)-\widehat{F}_{y_1}(F_{y_1}^{-1}(y)+\varepsilon)>C_1\varepsilon \text{ or } \widehat{F}_{y_1}(F_{y_1}^{-1}(y)-\varepsilon)-F_{y_1}(F_{y_1}^{-1}(y)-\varepsilon)>C_1\varepsilon\right) \\
\leq & \ \mathbb{P}\left(\left\|\widehat{F}_{y_1}-F_{y_1}\right\|>C_1\varepsilon\right).
\end{aligned}
$$

Because $\widehat{F}_{y_1}(.)$ converges uniformly to $F_{y_1}(.)$, (6.9) holds.

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

$$
P\left(\left\|\widehat{F}_{y_k}-F_{y_k}\right\|>\zeta\right)\leq\varepsilon/2. \tag{6.11}
$$

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

$$
\begin{aligned}
|\widehat{H}_{kj}(x)-H_{kj}(x)| & \leq |\widehat{F}_{y_j}^{-1}(\widehat{F}_{y_k}(x))-F_{y_j}^{-1}(\widehat{F}_{y_k}(x))|+|F_{y_j}^{-1}(\widehat{F}_{y_k}(x))-F_{y_j}^{-1}(F_{y_k}(x))| \\
& \leq \sup_{u\in K}|\widehat{F}_{y_j}^{-1}(u)-F_{y_j}^{-1}(u)|+C_2\left\|\widehat{F}_{y_k}-F_{y_k}\right\|,
\end{aligned} \tag{6.12}
$$

where $C_2=\sup_{u\in K}F_{y_j}^{-1\prime}(u)<\infty$ by (6.10). Fix $\delta>0$. By uniform convergence of $\widehat{F}_{y_1}(.)$ and (6.9), for all N large enough,

$$
P\left(\sup_{u\in K}|\widehat{F}_{y_j}^{-1}(u)-F_{y_j}^{-1}(u)|+C_2\left\|\widehat{F}_{y_k}-F_{y_k}\right\|>\delta\right)<\frac{\varepsilon}{2}. \tag{6.13}
$$

Then, for all N large enough,

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

where the second inequality stems from (6.11) and (6.12), and the third from (6.13). The result follows since ε and δ were arbitrary.

3. Consistency of $\widehat{y}_n$, for all $n\in\mathbb{Z}$.

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

$$
P(|\widehat{y}_n-y_n|\leq\varepsilon)\to 1 \tag{6.14}
$$

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

$$P(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \leq \varepsilon) \to 1 \tag{6.15}$$

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

$$H'_{12}(x) = \frac{f_{y_1}(x)}{f_{y_2}\left(F_{y_1}(x)\right)},$$

it follows, by (6.10), that $C_3 = 1 \vee \sup_{x \in B(y_{n-1}, \varepsilon)} |H'_{12}|(x) < \infty$. Moreover, by the induction hypothesis and the uniform convergence of $\widehat{H}_{12}(.)$, for all N large enough, the event

$$E_0 = \left\{ |\widehat{y}_{n-1} - y_{n-1}| < \varepsilon/2C_3, \sup_{x \in B(y_{n-1}, \varepsilon)} |\widehat{H}_{12}(x) - H_{12}(x)| < \varepsilon/2 \right\}$$

holds with an arbitrarily large probability. Under E_0,

$$\begin{aligned}
|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| &\leq |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(\widehat{y}_{n-1})| + |H_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \\
&\leq \sup_{x \in B(y_{n-1}, \varepsilon)} |\widehat{H}_{12}(x) - H_{12}(x)| + C_3 |\widehat{y}_{n-1} - y_{n-1}| \\
&\leq \varepsilon.
\end{aligned}$$

This proves (6.15) and concludes the induction step. Thus, (6.14) holds for all $n \in \mathbb{Z}$.

4. Consistency of the estimated bounds of $C'(.)$ and $F_\theta(.)$.

We focus on $\overline{C'}(.)$ and $\overline{F}_\theta(.)$, the reasoning being similar for the lower bounds. Let $\widehat{\underline{\theta}}_1(y) = \sup_{n \in \mathbb{Z}:\, \widehat{y}_n \geq y} \theta_n$ and $\widehat{\underline{y}}_1(\theta) = \sup_{n \in \mathbb{Z}:\, \theta_n \geq \theta} \widehat{y}_n$. By definition of $\widehat{\overline{C'}}(.)$, it suffices to prove that $\widehat{\underline{\theta}}_1(.)$ is consistent for all $y \notin \{y_n, n \in \mathbb{Z}\}$. Similarly, by definition of $\widehat{\overline{F_\theta}}(.)$ and uniform consistency of $\widehat{F}_{y_k}(.)$, it suffices to prove that $\widehat{\underline{y}}_1(.)$ is consistent for all $\theta > 0$.

Let us begin with $\widehat{\underline{\theta}}_1(y)$. Because $(\theta_n)_{n \in \mathbb{Z}}$ is decreasing,

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

where $\underline{n}_1(y) = \min\{n \in \mathbb{Z} :\, y_n \geq y\}$. Moreover, because $y \notin \{y_n, n \in \mathbb{Z}\}$,

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

Let us consider the event

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

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

We now turn to $\widehat{\underline{y}}_1(.)$. Because $(y_n)_{n\in\mathbb{Z}}$ is increasing, $\underline{y}_1(\theta) = y_{\overline{n}_1(\theta)}$, where $\overline{n}_1(\theta) = \max\{n \in \mathbb{Z} : \theta_n \geq \theta\}$. Similarly, $\widehat{\underline{y}}_1(\theta) = \widehat{y}_{\overline{n}_1(\theta)}$. By convergence of $\widehat{y}_{\overline{n}_1(\theta)}$, $\widehat{\underline{y}}_1(\theta)$ converges to $\underline{y}_1(\theta)$. ■