DISCUSSION PAPER SERIES

IZA DP No. 4606

# Inference on a Generalized Roy Model, with an Application to Schooling Decisions in France

Xavier d'Haultfoeuille
Arnaud Maurel

December 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Inference on a Generalized Roy Model, with an Application to Schooling Decisions in France


**Xavier d'Haultfoeuille**
*CREST-INSEE*

**Arnaud Maurel**
*CREST-INSEE, ENSAE, PSE and IZA*



Discussion Paper No. 4606
December 2009



IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4606
December 2009

# ABSTRACT

## Inference on a Generalized Roy Model, with an Application to Schooling Decisions in France*

This paper considers the identification and estimation of an extension of Roy's model (1951) of occupational choice, which includes a non-pecuniary component in the decision equation and allows for uncertainty on the potential outcomes. This framework is well suited to various economic contexts, including educational and sectoral choices, or migration decisions. We focus in particular on the identification of the non-pecuniary component under the condition that at least one variable affects the selection probability only through potential earnings, that is under the opposite of the usual exclusion restrictions used to identify switching regressions models and treatment effects. Point identification is achieved if such variables are continuous, while bounds are obtained otherwise. As a result, the distribution of the *ex ante* treatment effects can be point or set identified without any usual instruments. We propose a three-stages semiparametric estimation procedure for this model, which yields root-n consistent and asymptotically normal estimators. We apply our results to the educational context, by providing new evidence from French data that non-pecuniary factors are a key determinant of higher education attendance decisions.

JEL Classification: C14, C25, J24

Keywords: Roy model, nonparametric identification, exclusion restrictions, schooling choices, ex ante returns to schooling

Corresponding author:

Arnaud Maurel
ENSAE
Bureau E12 - Timbre J120
3, Avenue Pierre Larousse
92245 Malakoff cedex
France
E-mail: maurel@ensae.fr

* We thank Christian Belzil, Philippe Février, Bo Honoré, Simon Lee, Thierry Magnac, Jean-Marc Robin, Gerard van den Berg, participants at the annual conference of the ESRC Econometric Study Group (Bristol, July 2009), at the ESEM (Barcelona, August 2009), at the summer school Stats in the Château (Jouy-en-Josas, September 2009) and in seminars at CREST-INSEE (Paris, October 2009) and at Mannheim University (November 2009) for helpful discussions and comments at various stages of this research.

# 1 Introduction

Self-selection is probably one of the major issue economists have to deal with when trying to measure causal effects such as, among others, returns to education, returns to sectoral choice as well as migration benefits. The seminal Roy's model (1951) of occupational choice can be seen as an extreme setting of self-selection, where agents choose the sector which provides them with the higher wage. The idea underlying this model has been very influential in the analysis of choices of participation to the labor market (Heckman, 1974), union versus nonunion status (Lee, 1978, Robinson & Tomes, 1984), public versus private sector (Dustmann & van Soest, 1998), college attendance (Willis & Rosen, 1979), migration (Borjas, 1987), training program participation (Ashenfelter & Card, 1985, Ham & LaLonde, 1996) as well as occupation (Dolton et al., 1989).

The standard Roy model is, however, restrictive in at least two dimensions. First, non-pecuniary aspects matter much in general. For instance, in the context of educational choice, it is most often assumed that individuals consider not only the investment value of schooling, which is related to wage returns, but also the non-pecuniary consumption value of schooling, which is related to preferences and schooling ability. Recent empirical evidence suggest that these non-pecuniary factors are indeed a key determinant of schooling decisions (Carneiro et al., 2003, and Beffy et al., 2009). Non-pecuniary aspects such as working conditions may also matter when choosing an occupation. Similarly, migration decisions are likely to be driven both by the *ex ante* monetary returns and the psychic costs associated with the decision to migrate (Bayer et al., 2008). Second, as emphasized by a recent stream of the literature on schooling choice (see Cunha & Heckman, 2007, for a survey), agents most often do not anticipate perfectly their potential earnings in each sector at the moment of their decision. Because of *ex ante* uncertainty, their decision depends on expectations of these potential earnings rather than on their true values.[1]

In this paper, we explore what can be nonparametrically identified in a generalized Roy model including these two aspects, when relying extensively on its detailed structure. An original feature of our approach lies in the fact that we do not need any standard instrument, that is we do not rely on the availability of a variable affecting the selection probability but not the potential earnings. Such instruments do not exist, for instance, when the true model is a standard Roy model, possibly extended to account for *ex ante* uncertainty. We first develop two strategies for identifying the covariates effects on sector-

[1]Note that even if there is no *ex ante* uncertainty, the decision also depends on expectation of the measured outcomes if they are affected by (standard) measurement errors.

specific earnings. The first one is based on exclusion restrictions between sector-specific regressors, while the second one exploits an argument at infinity, which relies on a recent result from a companion paper (d'Haultfoeuille & Maurel, 2009). We then study identification of the non-pecuniary component under the condition that at least one regressor affects the selection probability only through *ex ante* monetary returns. By doing so, we extend the standard Roy model, which does not include any non-pecuniary component, to a setting where this component only varies according to a subset of the regressors. Although natural, this kind of identifying condition has received very little attention in the literature. d'Haultfoeuille (2010) considers a similar condition in sample selection models but his assumption breaks down in the model considered here because of *ex ante* uncertainty in the potential outcomes. Carneiro et al. (2003) also exploit instruments of these kinds to estimate an extension of the Willis and Rosen's model (1979) of demand for college attendance, but without considering their identifying power. Under this assumption, we show that the non-pecuniary component is point identified when at least one instrument is continuous. When the instruments are discrete, we provide easy to compute bounds on this non-pecuniary component. Noteworthy, our results are not based neither on a large support condition on the covariates nor on parametric restrictions.

Finally, we show that the identification of the covariates effects and the non-pecuniary component conveys information about the distribution of causal treatment effects. Even if no standard instrument is available, we obtain bounds on the distribution of the monetary benefits anticipated by the agents, which correspond in this setting to the marginal treatment effect (see Heckman & Vytlacil, 2005). Standard average treatment effects are point identified if the probability of selection ranges from zero to one, a result in line with the one of Heckman & Vytlacil (2005) in the case of standard instrumental variable strategies.

On a related ground, a recent paper by Bayer et al. (2008) also considers the identification of a generalized Roy model accounting for non-pecuniary factors. Our approach differs from theirs in two main aspects. First, Bayer et al. (2008) do not account for *ex ante* uncertainty, which may often be large. Second, their identification results are obtained under alternative assumptions. They first show that the non-pecuniary factors associated with each choice alternative and the unconditional wage distributions are identified provided that the distribution of pecuniary returns has a finite lower bound. Although appealing in that it does not require any exclusion restriction, the finite support assumption may be restrictive, in particular when using log wages in utility functions, as for instance in Willis & Rosen (1979). Bayer et al. (2008) alternatively prove identification under the assumption of independence between alternative-specific wages and the exclusion restriction that a variable affects the non-pecuniary valuation of each choice alternative but not the wage distributions.[2] The independence assumption, however, is restrictive, and much of the literature considering identification of Roy and the closely related competing risks models has produced alternative identification results without such an assumption (see, e.g., Heckman & Honore, 1989, Heckman & Honore, 1990 or Abbring & van den Berg, 2003). The identification results we derive in our paper do not rely neither on the aforementioned support condition nor on that independence assumption.

Apart from identification, we also propose a three-stages semiparametric estimation procedure when the effects of the covariates are linear. The first two stages allow to estimate the covariates effects on potential earnings and correspond to Newey's method (2008) for estimating semiparametric selection models. The originality of the proposed estimation procedure lies in its third stage, which is devoted to the non-pecuniary component. This stage is rather simple as it amounts to estimate an instrumental linear model. The only difficulty lies in estimating the dependent variable of this linear model, as it involves both the first steps estimators and a nonparametric nuisance parameter. We show that the corresponding estimator is root-n consistent and asymptotically normal. Monte Carlo simulations indicate that despite its multiple steps, the estimators perform fairly well in finite samples.

Eventually, in the empirical section of the paper, we apply our semiparametric estimation procedure to the context of higher education attendance decisions in France over the nineties. For that purpose we suppose, in a same spirit as Carneiro et al. (2003), that the local average income for high school graduates only affects the probability of attendance through the *ex ante* returns to higher education. Consistently with the recent empirical evidence on this question, our results suggest that non-pecuniary factors are a key determinant of the decision to attend higher education. We are able to compare the influence of non-pecuniary factors with the one of *ex ante* monetary returns to education, the distribution of these returns being point identified on most of its support. Noteworthy, unlike Carneiro et al. (2003), our results are obtained without imposing a factor structure on the outcomes. According to our estimates, the median of these factors in the population represents 2.5 times the median of the returns to higher education, thus highlighting the major role played by non-pecuniary determinants in the decision to enroll in higher education.

The remainder of the paper is organized as follows. Section 2 presents the extended Roy model which is considered throughout the paper and gives identification results for the covariates effects on earnings and for the non-pecuniary component. Section 3 develops a

[2]Bayer et al. (2008) refer to this exclusion restriction as the Commonality assumption.

semiparametric estimation procedure for the extended Roy model, and proves root-n consistency and asymptotic normality of the proposed estimators. Section 4 studies finite-sample performances of the estimators. Section 5 applies the preceding estimators to recover an estimate of the influence of non-pecuniary factors on higher education attendance decision in France. Finally, Section 6 concludes. The proofs of our results are deferred to Appendix A.

# 2 Identification

## 2.1 The setting

We consider an extension of the Roy model which is obtained by including *ex ante* uncertainty as well non-pecuniary factors in the seminal Roy's model (1951) of occupational choice. Suppose that there are two sectors 0 and 1 in the economy, and let $Y_k,\ k \in \{0, 1\}$, denote the individual's potential outcome in sector $k$.[3] These outcomes are not perfectly observed by the individual at the time of her decision. Instead, she can only compute the expectation $E(Y_k|X, \eta_0, \eta_1)$, where $X$ are covariates observed by the econometrician and $(\eta_0, \eta_1)$ are sector-specific productivity terms known by the agent at the time of the choice but unobserved by the econometrician. We will maintain the following assumption throughout the article.

**Assumption 2.1** *(Additive decomposition) We have, for $k \in \{0, 1\}$, $E(Y_k|X, \eta_0, \eta_1) = E(Y_k|X, \eta_k) = \psi_k(X) + \eta_k$. Moreover, $X \perp\!\!\!\perp (\eta_0, \eta_1)$.*

We can always suppose that $\eta_k$ is mean independent of $X$, i.e. $E(\eta_k|X)$ is constant. We reinforce here mean independence into independence, ruling out for instance heteroskedasticity. Such an assumption is commonly made when studying sample selection models (see, e.g., Newey, 2008) or the standard Roy model (see, e.g., Heckman and Honoré, 1990). Besides, we let $\nu_k = Y_k - E(Y_k|X, \eta_0, \eta_1)$ denote the unexpected shock on $Y_k$ and $\varepsilon_k = \eta_k + \nu_k$ denote the sector-specific residual. Note that apart from the independence assumption, we do not impose any restriction on $(\eta_0, \eta_1, \nu_0, \nu_1)$, thus departing from, e.g., Carneiro et al. (2003).[4]

[3] Notice that the subscript $k$ refers to the sector and not to the individual. For the sake of simplicity and in the absence of ambiguity, individual subscripts are omitted in this section.

[4] Carneiro et al. (2003) impose in particular a factor structure on the unobservables. Such restrictions are useful to identify the joint distribution of $(\eta_0, \eta_1, \nu_0, \nu_1)$, and thus to test for comparative advantage or

Unlike Roy's original model, we do not suppose that the sectoral choice is based only on income maximization. Instead, we suppose that each individual chooses to enter the sector which yields the highest expected utility, with the expected utility in sector $k$ writing as $\mathcal{U}_k = E(Y_k|X, \eta_0, \eta_1) + G_k(X)$. Hence, $\mathcal{U}_k$ is assumed to be given by the sum of sector-specific expected outcome $E(Y_k|X, \eta_0, \eta_1)$ and the non-pecuniary component associated with sector $k$, $G_k(X)$, which is supposed to depend on the covariates $X$. Thus, along with the covariates $X$, the econometrician observes the chosen sector $D$, which satisfies

$$\begin{aligned} D &= \mathbb{1}\{\mathcal{U}_1 > \mathcal{U}_0\} \\ &= \mathbb{1}\{\eta_\Delta > \psi_0(X) - \psi_1(X) + G(X)\}, \end{aligned} \tag{2.1}$$

where $G(X) = (G_0 - G_1)(X)$ and $\eta_\Delta = \eta_1 - \eta_0$. Finally, the econometrician also observes the outcome in the chosen sector, that is

$$Y = DY_1 + (1 - D)Y_0.$$

This model is quite general and can be applied to various economic settings, including sectoral choice in the labor market, immigration or higher education attendance decisions (see our application in Section 5). It is close to the class of generalized Roy models which are considered in the treatment effects literature (see e.g. Heckman & Vytlacil, 2005).[5] The difference lies in the fact that in these models, the factor $G$ is random and can be correlated with $(\eta_0, \eta_1, \nu_0, \nu_1)$ in an unspecified way. Imposing our structure has two main advantages with respect to the treatment effects literature. First, we are able to recover the non-pecuniary factors entering the selection equation, and compare them with the *ex ante* monetary returns which correspond in this setting to the marginal treatment effect. Second, our approach does not rely on an instrument that affects the selection but not the potential outcomes. We rely on the alternative condition that at least one regressor affects the selection probability only through potential outcomes. In some contexts, this kind of exclusion restriction may actually be easier to find (see our discussion in Subsection 2.3).

We will maintain the following assumptions subsequently.

**Assumption 2.2** *(Normalization) There exists $x^*$ such that $\psi_0(x^*) = \psi_1(x^*) = 0$.*

to assess the importance of *ex post* uncertainty (see Cunha & Heckman, 2007). We do not consider these issues here.

[5] We refer here to the static treatment effects literature. See the extension by Heckman & Navarro (2007), who consider identification of dynamic discrete choice models which are used as underlying structural frameworks for dynamic treatment effects.

**Assumption 2.3** *(Restrictions on the errors, 1)* $E(|\varepsilon_k|) < \infty$ *for* $k \in \{0, 1\}$*. Moreover, the distribution of* $\eta_\Delta$ *admits a density, denoted by* $f_{\eta_\Delta}$*, with respect to the Lebesgue measure.*

Assumption 2.2 is an innocuous normalization which stems from the fact that adding a constant to $\psi_k$ and subtracting it to $\eta_k$ does not modify the model. Assumption 2.3 is a technical condition which is usual in competing risks or Roy models (see in particular Heckman & Honore, 1990, for the case of Roy model and Lee, 2006 for the case of competing risks models.).

### 2.2 Identification of $(\psi_0, \psi_1)$

Before detailing our key result on the identification of $G$, we present in this subsection two strategies to recover $(\psi_0, \psi_1)$. The first is rather standard and relies on exclusion restrictions, in a similar spirit as in, e.g., Heckman & Honore (1990). The second yields identification at infinity, and presents the advantage of not requiring any exclusion restriction. The first strategy relies on the following assumption.

**Assumption 2.4** *(Exclusion restrictions, 1) There exists* $X_0, X_1, X_c$ *such that* $X = (X_0, X_1, X_c)$ *and* $\psi_0$ *(resp.* $\psi_1$*) depends only on* $(X_0, X_c)$ *(resp. on* $(X_1, X_c)$*). Moreover,* $(X_0, X_c)$ *(resp.* $(X_1, X_c)$*) and* $P(D = 1|X)$ *are measurably separated, that is, any function of* $(X_0, X_c)$ *(resp. of* $(X_1, X_c)$*) almost surely equal to a function of* $P(D = 1|X)$ *is almost surely constant.*

Basically, the measurable separation requirement[6] of Assumption 2.4 ensures that $\psi_0(X)$ (or $\psi_1(X)$) and $P(D = 1|X)$ can vary in a sufficiently independent way. This assumption is weak and is, for instance, assumed implicitly in nonparametric additive regression (see, e.g., Linton & Nielsen, 1995). The first part of Assumption 2.4 covers two rather different situations. The first one is when $X_0 = X_1 = \emptyset$ but we observe some variables which affect the non-pecuniary component but not the potential outcomes. This situation corresponds to the standard instrumental setting in sample selection models as well as to the commonality condition of Bayer et al. (2008). The other one is when we observe some variables $X_0$ and $X_1$ which affect only one sector. In this latter case, no exclusion restriction between the non-pecuniary factors and the potential outcomes is required.

Given the preceding exclusion restrictions and the additive decomposition assumption, it is possible to identify $\psi_0$ and $\psi_1$ up to location parameters. Then full identification stems

[6] We adopt here the terminology of Florens et al. (2008) (see their Assumption A4).

from the normalization of Assumption 2.2. Note that Theorem 2.1 does not provide any result on the location parameters. In general, such parameters are identified only at infinity, i.e. when $P(D=1|X)$ can be arbitrarily close to zero and one (see, e.g., Heckman, 1990).

**Theorem 2.1** *Suppose that Assumptions 2.1-2.4 hold. Then $\psi_0$ and $\psi_1$ are identified.*

Alternatively, $\psi_0$ and $\psi_1$ can also be identified at the limit without any exclusion restriction, under the following restrictions on the error terms.

**Assumption 2.5** *(Restrictions on the errors, 2) (i) $X \perp\!\!\!\perp (\varepsilon_0, \varepsilon_1)$, (ii) for $k \in \{0,1\}$, the supremum of the support of $\varepsilon_k$ is infinite and there exists $b_k > 0$ such that $E(\exp(b_k \varepsilon_k)) < \infty$, (iii) for all $u \in \mathbb{R}$,*

$$\lim_{v \to \infty} P(\eta_k - \eta_{1-k} > u | \eta_k + \nu_k = v) = 1, \quad k \in \{0,1\}.$$

The first restriction reinforces the condition that $X \perp\!\!\!\perp (\eta_0, \eta_1)$, by ruling out in particular heteroskedasticity on the shocks $(\nu_0, \nu_1)$. The second restriction is a light tail condition, which is in practice fairly mild.[7] The last one can be interpreted as a moderate dependence condition between $\eta_0$ and $\eta_1$. When $(\eta_0, \eta_1, \nu_0, \nu_1)$ is gaussian for instance, one can show that it is equivalent to $\text{cov}(\eta_0, \eta_1) < \min(V(\eta_0), V(\eta_1))$. In particular, when $V(\eta_0) = V(\eta_1)$, this condition is automatically satisfied, except in the degenerated case where $\eta_0 = \eta_1$.

**Theorem 2.2** *Suppose that Assumptions 2.1, 2.2 and 2.5 hold. Then $\psi_0$ and $\psi_1$ are identified.*

Theorem 2.2 is based on a result by d'Haultfoeuille & Maurel (2009), and on the fact that under Assumption 2.5,

$$\lim_{y \to \infty} P(D = k | X = x, Y_k = y) = 1, \text{ for all } x \text{ and } k \in \{0,1\}. \tag{2.2}$$

In other words, individuals whose potential outcome in one sector tends to infinity will choose this sector with a probability approaching one. Intuitively, this condition implies that there is no selection issue when one of the potential outcome becomes arbitrarily large. The idea of identification at infinity is similar to the one obtained by Heckman & Honore (1989) and Abbring & van den Berg (2003) in the related competing risks model. Their

[7] If we consider the example of log-wages $Y_k = \ln W_k$, the assumption is satisfied provided that there exists $b_k > 0$ such that $E(W_k^{b_k}) < \infty$. Hence, it holds even if wages have fat tails, Pareto like for instance.

results can nevertheless not be used here because their strategies break down when turning to generalized Roy models.[8]

An appealing feature of Condition (2.2) is that it is testable (see d'Haultfoeuille & Maurel, 2009). On the other hand, estimators corresponding to this setting have not been derived yet. Therefore, we restrict in the estimation part (Section 3) to the case where exclusion restrictions are available.

### 2.3 Identification of the non-pecuniary component

We now turn to the identification of $G$. We will suppose for that purpose that one of the two frameworks displayed above can be used to identify $(\psi_0, \psi_1)$. Then $\varepsilon = Y - \psi_D(X)$ and $T = \psi_0(X) - \psi_1(X)$ are identified. Let $X = (W, Z)$. Our analysis relies on the following exclusion restriction.

**Assumption 2.6** *(Exclusion restrictions, 2) Almost surely, $G$ only depends on $W$. Moreover, the distribution of $T$ conditional on $W$ is not degenerated.*

Assumption 2.6 allows us to make $T$ vary while holding the non-pecuniary component $G(X)$ fixed. Hence, for Assumption 2.6 to be verified, one needs a variable which determines the sector-specific potential outcomes but does not enter the non-pecuniary component. This is the opposite of the kind of exclusion restrictions which are most often used to identify labor supply and more generally switching regressions models. In some empirical situations, one may feel more comfortable with the use of exclusion restrictions of that kind. In the example of college attendance decision, standard instruments which are assumed to affect earnings only indirectly through college attendance include in particular parental background and distance to college. These instruments have been criticized on various grounds, related in particular to the intergenerational transmission of ability and to the endogeneity of geographical mobility.[9] We propose to use instead local labor market conditions such as average local labor market income. Noteworthy, though primarily relying on standard exclusion restrictions to identify their model of college attendance, Carneiro et al. (2003) also exploit similar exclusion restrictions. This kind of exclusion restrictions involving local labor market income or local unemployment rate is actually quite natural in many

[8] Lee (2006) and Lee & Lewbel (2009) obtain identification of competing risks models without using arguments at the limit. Their strategy cannot be extended easily either to generalized Roy models.

[9] It is also standard in this literature to exploit variations in attendance rates induced by the level of tuition fees. Nevertheless, as detailed further, tuition fees are very low in France and vary little across regions, so that they cannot be used as an instrument for higher education attendance.

other economic situations, including the decision to apply for public sector jobs as well as migration decisions.

Under Assumption 2.6, $G$ can be identified up to a location parameter, using only the facts that $T$ is identified and $P(D=0|X) = F_{\eta_\Delta}(T+G(X))$, where $F_{\eta_\Delta}$ denotes the distribution function of $\eta_\Delta$. Such an identification can be achieved for instance if $(i)$ there exists $w_1$ such that for all $w$ in the support of $W$, there exists $z, z_1$ such that $P(D=1|W=w, Z=z) = P(D=1|W=w_1, Z=z_1)$, and $(ii)$ $F_{\eta_\Delta}$ is strictly increasing. However, this result is not as positive as it might seem. Aside from being obviously necessary to assess the weight of non-pecuniary factors, the location of $G$ is indeed also crucial to determine the distribution of treatment effects. For instance, the distribution function $F_\Delta$ of the *ex ante* treatment effect $\Delta = E(Y_1 - Y_0|X, \eta_0, \eta_1)$ satisfies

$$\begin{aligned} F_\Delta(u) &= P(E(Y_1|X,\eta_0,\eta_1) - E(Y_0|X,\eta_0,\eta_1) \leq u) \\ &= E\left(F_{\eta_\Delta}(u+T)\right). \end{aligned} \tag{2.3}$$

Besides, recall that the selection equation implies that

$$P(D=0|X) = F_{\eta_\Delta}(T+G(X)). \tag{2.4}$$

If $G$ is identified only up to a location parameter, then one can shift $F_{\eta_\Delta}$ by any real number, thus implying that in general this identification result does not yield informative bounds on $F_\Delta$.

We now show that the detailed structure of the generalized Roy model actually provides either full or partial identification of the whole non-pecuniary component, including its location. In the following, we omit the dependence in $W$ for the ease of notation. Thus the results must be understood to be conditional on $W$. We start from the following observations:

$$\begin{aligned} E[D\eta_\Delta|T=t] &= E\left[\mathbb{1}\{\eta_\Delta \geq t+G\}\eta_\Delta\right] = \int_{t+G}^{\infty} u f_{\eta_\Delta}(u)du, \\ E[D|T=t] &= \int_{t+G}^{\infty} f_{\eta_\Delta}(u)du. \end{aligned}$$

First, suppose that $T$ is continuous. Then, letting $q_0(t) = E(D|T=t)$, we obtain

$$\frac{\partial E[D\eta_\Delta|T=t]}{\partial t} = (t+G)q_0'(t).$$

Now, the definition of $\nu_i$ and the law of iterated expectations yield $E(\nu_i|D=i,T)=0$. As a result,

$$\begin{aligned} E(\varepsilon|T=t) &= E[D\varepsilon_1 + (1-D)\varepsilon_0|T=t] \\ &= E[D\eta_1 + (1-D)\eta_0|T=t] \\ &= E\left[D\eta_\Delta|T=t\right] + E[\eta_0]. \end{aligned}$$

Thus, letting $g_0(t) = E(\varepsilon|T = t)$, we get

$$g_0'(t) = (t + G)q_0'(t). \tag{2.5}$$

This equation ensures the identification of $G$ provided that $q_0'(t) \neq 0$ for at least one $t \in \mathcal{S}$, where $\mathcal{S}$ denotes the support of $T$. This will be the case for instance if $P(\eta_\Delta - t \in \mathcal{S}) > 0$ for all $t \in \mathbb{R}$, or under the stronger condition that $f_{\eta_\Delta}(u) > 0$ for all $u \in \mathbb{R}$. For the matter of convenience, we suppose subsequently that the latter condition holds.

**Assumption 2.7** *For all $u \in \mathbb{R}$, $f_{\eta_\Delta}(u) > 0$.*

Now consider the case where $T$ has a discrete distribution and takes $M$ values $t_1 < t_2 < ... < t_M$. Then we cannot take the derivative of $g_0$ and $q_0$ anymore. However, the strategy above can be adapted to yield bounds on $G$. Indeed, letting $i < j$, we have,

$$\begin{aligned}
& \sum_{k=i}^{j-1} t_{i+1}(q_0(t_{i+1}) - q_0(t_i)) + G(q_0(t_j) - q_0(t_i)) \\
\leq \; & g_0(t_j) - g_0(t_i) = -\int_{t_i+G}^{t_j+G} u f_{\eta_\Delta}(u)du \\
\leq \; & \sum_{k=i}^{j-1} t_i(q_0(t_{i+1}) - q_0(t_i)) + G(q_0(t_j) - q_0(t_i)).
\end{aligned}$$

In other words, $G \in [\underline{G}_{ij}, \overline{G}_{ij}]$ with

$$\begin{aligned}
\underline{G}_{ij} &= \frac{\sum_{k=i}^{j-1} t_{i+1}(q_0(t_{i+1}) - q_0(t_i)) + g_0(t_i) - g_0(t_j)}{q_0(t_i) - q_0(t_j)} \\
\overline{G}_{ij} &= \frac{\sum_{k=i}^{j-1} t_i(q_0(t_{i+1}) - q_0(t_i)) + g_0(t_i) - g_0(t_j)}{q_0(t_i) - q_0(t_j)}.
\end{aligned}$$

Note that these ratios are well defined since Assumption 2.7 ensures that $q_0(t_i) > q_0(t_j)$. Finally, we can improve these bounds by optimizing over $i < j$. We sum up our results in the following theorem.

**Theorem 2.3** *Suppose that $(\psi_0, \psi_1)$ are identified, and that Assumptions 2.1, 2.3, 2.6 and 2.7 hold. Then:*

- *if the distribution of $T$ is continuous, $G$ is identified;*
- *if the distribution of $T$ is discrete and takes values in $\{t_1 < t_2 < ... < t_M\}$, then $G \in [\underline{G}, \overline{G}]$, with $\underline{G} = \max_{i<j} \underline{G}_{ij}$ and $\overline{G} = \min_{i<j} \overline{G}_{ij}$.*

In the discrete case, the length of the interval is lower than $\min_{i<j} t_j - t_i$. This can be best seen when $M = 2$, where it is actually equal to $t_2 - t_1$. Hence, contrary to many other examples in econometrics, large variations in the data are not desirable for identifying $G$. On the other hand, such large variations may improve the accuracy of the related estimators, since when $t_i$ is close to $t_j$, $q_0(t_j) - q_0(t_i)$ is close to zero and the fluctuations of the estimated denominator of $\underline{G}_{ij}$ and $\overline{G}_{ij}$ are likely to make the estimators more unstable.

In the continuous case, identification of $G$ can be achieved through Equation (2.5). However, this equation involves derivatives of nonparametric regressions, which are not estimated accurately. Integrating between $t_0 \in \mathcal{S}$ and $T$ and using an integration by part, we obtain

$$g_0(T) - Tq_0(T) + \int_{t_0}^{T} q_0(u)du = \alpha_0 + Gq_0(T), \tag{2.6}$$

where $\alpha_0 = g_0(t_0) - t_0 q_0(t_0) - Gq_0(t_0)$. In other words,

$$\varepsilon - DT + \int_{t_0}^{T} q_0(u)du = \alpha_0 + DG + \xi, \quad E(\xi|T) = 0 \tag{2.7}$$

This equation is more convenient for estimating $G$ as it does not depend on derivatives terms. Moreover, once the left term has been estimated nonparametrically, it reduces to a linear instrumental equation with only one regressor.[10]

### 2.4 Distribution of treatment effects

We now turn to the identification of the distribution of the *ex ante* treatment effect, $\Delta = E(Y_1 - Y_0|X, \eta_0, \eta_1)$. *Ex ante* treatment effects are meaningful since they correspond to what agents act on (see Cunha & Heckman, 2007). Besides, it corresponds to the *ex post* treatment effect if (i) agents perfectly observe their potential outcomes (in which case $\nu_0 = \nu_1 = 0$) or if (ii) the idiosyncratic shocks are equal across sectors ($\nu_0 = \nu_1$), as postulated in standard regression models.[11] Having identified $T$ and $G(.)$, the selection Equation (2.4) shows that $F_{\eta_\Delta}$ is identified for all $u$ in the support of $T + G(X)$. Thus, by Equation (2.3), one can identify $F_\Delta(u)$ for all $u$ such that the support of $u+T$ is included in the support of $T + G(X)$. In particular, the complete distribution of the *ex ante* treatment

[10] In the case where the distribution of $T$ conditional on $W$ is discrete with an infinite support, Equation (2.7) can be used to obtain bounds on $G$. In this case $q_0(u)$, defined as $S_{\eta_\Delta}(u + G)$ (where $S_{\eta_\Delta}$ denotes the survival function of $\eta_\Delta$), is observed only for $u$ in the support of $T$, so that the integral term is not point identified in general. However, the monotonicity of $q_0(.)$ allows to bound this integral term, and one may apply for instance Manski & Tamer's (2002) results to set identify $G$.

[11] We conjecture that without further assumption on $(\nu_0, \nu_1)$, not much can be learned on the distribution of $Y_1 - Y_0$.

effects $\Delta$ will be identified as soon as $T+G(X)$ has a large support. In that case, one can recover standard treatment effect parameters such as the average treatment effect or the average treatment on the treated. But even if this large support condition fails, it is still possible to point identify a subset of the distribution of the *ex ante* treatment effect, and bound $F_\Delta(u)$ for the rest of the distribution. Indeed, letting $[\underline{M},\overline{M}]$ (resp. $[\underline{P},\overline{P}]$) denote the support of $T+G(X)$ (resp. of $P(D=0|X)$), we have, by the monotonicity of $F_{\eta_\Delta}$, $F_\Delta(u)\in[\underline{F_\Delta}(u),\overline{F_\Delta}(u)]$, with

$$\begin{aligned}\underline{F_\Delta}(u) &= E\left(F_{\eta_\Delta}(u+T)\,\mathbb{1}\{u+T\in[\underline{M},\overline{M}]\}\right)\\ &\quad +\overline{P}\times P(u+T>\overline{M})+0\times P(u+T\le\underline{M}) \qquad (2.8)\\ \overline{F_\Delta}(u) &= E\left(F_{\eta_\Delta}(u+T)\,\mathbb{1}\{u+T\in[\underline{M},\overline{M}]\}\right)\\ &\quad +1\times P(u+T>\overline{M})+\underline{P}\times P(u+T\le\underline{M}). \qquad (2.9)\end{aligned}$$

The distribution of the *ex ante* treatment effects on the treated can be identified in a similar way, with

$$F_{\Delta|D=1}(u)=\frac{E\{(F_{\eta_\Delta}(u+T)-P(D=0|X))\times\mathbb{1}\{G(X)\le u\}\}}{P(D=1)}. \qquad (2.10)$$

In our setting, the *ex ante* treatment effect $\Delta$ is closely related to the marginal treatment effect $\Delta^{MTE}$ (Heckman & Vytlacil (2005)). Indeed, denoting by $S_{\eta_\Delta}$ the survival function of $\eta_\Delta$, we have, under Assumption 2.7,

$$\Delta^{MTE}(x,p)=\psi_1(x)-\psi_0(x)+S^{-1}_{\eta_\Delta}(p)$$

Thus, $\Delta=(\psi_1-\psi_0)(X)+\eta_\Delta$ coincides with $\Delta^{MTE}(X,S_{\eta_\Delta}(\eta_\Delta))$. Besides, one is able to identify $\Delta^{MTE}(x,p)$ for all $p$ in the support of $P(D=1|X)$, since in that case there exists $x$ in the support of $X$ such that $S^{-1}_{\eta_\Delta}(p)=(\psi_0-\psi_1+G)(x)$. Hence, the generalized Roy model we consider allows to identify $\Delta^{MTE}(x,.)$ on an interval which is generally larger than with the local instrumental variable approach considered by Heckman & Vytlacil (2005).[12]

## 3 Semiparametric estimation

Although our identification results hold in a nonparametric setting, we focus here on semiparametric estimation in order to provide root-$n$ consistent and asymptotically normal

[12] In this latter case indeed, $\Delta^{MTE}(x,.)$ is identified only on the support of $P(D=1|X=x,Z)$, where $Z$ denotes a regressor affecting $D$ which is excluded from the outcome equations. Intuitively, by allowing to make all the regressors vary, our approach provides identification of $\Delta^{MTE}(x,.)$ on a wider interval.

estimators of $\psi_0(.), \psi_1(.)$ and $G(.)$. More precisely, we consider generalized Roy models with linear index structures of the form:[13]

$$\begin{cases} Y_0 &=& X'\beta_0 + \varepsilon_0 \\ Y_1 &=& X'\beta_1 + \varepsilon_1 \\ D &=& \mathbb{1}\{-\delta_0 + X'(\beta_1 - \beta_0 - \gamma_0) + \eta_\Delta > 0\}. \end{cases} \tag{3.1}$$

In this setting, the non-pecuniary component is of the form $\delta_0 + X'\gamma_0$. Let $\gamma_{0k}$ (resp. $\beta_{0k}$, $\beta_{1k}$) denote the $k$-th component of $\gamma_0$ (resp. $\beta_0$, $\beta_1$), and let $\lambda_0 = \beta_0 - \beta_1 + \gamma_0$. We impose the following assumptions, which correspond to the exclusion restrictions of Assumptions 2.4 and 2.6, as well as to the continuity of $T$ conditional on $W$.

**Assumption 3.1** *(Exclusion restrictions) $\beta_{01} = \beta_{12} = 0$, $\gamma_{01} \neq \beta_{11}$ and $\gamma_{02} \neq -\beta_{02}$. Moreover, there exists $m$ such that $\gamma_{0m} = 0$ and $\beta_{0m} \neq \beta_{1m}$.*

**Assumption 3.2** *(Regularity of $X$) The support of $X$ is bounded. For all $x_{-m}$ in the support of $X_{-m} = (X_1, ..., X_{m-1}, X_{m+1}, ...)$, the distribution of $X_m$ conditional on $X_{-m} = x_{-m}$ admits a continuously differentiable and positive density on its support, which is a compact interval independent of $x_{-m}$. Moreover, for all $j$, $t \mapsto E(X_j|X'\lambda_0 = t)$ is continuously differentiable. Finally, the support of $X'\lambda_0$ is an interval.*

Assumption 3.2 ensures that there is at least one continuous instrument, namely $X_m$. As shown by Theorem 2.3, this condition is sufficient to provide point identification of $G(X)$. We also impose the support of $X'\lambda_0$ to be an interval. This condition is needed in general to obtain point identification in single index models (see, e.g., Horowitz, 1998). We propose a three-stages estimation procedure of the preceding model based on a sample of $(Y = DY_1 + (1 - D)Y_0, X, D)$.

**Assumption 3.3** *(i.i.d. sample) We observe a sample $(Y_i, X_i, D_i)_{1\leq i\leq n}$ of i.i.d. copies of $(Y, X, D)$.*

Let us assume, without loss of generality, that $\beta_{1m} - \beta_{0m}$ is strictly positive. We define $\zeta_0 = -\lambda_0/(\beta_{1m} - \beta_{0m})$, so that $\zeta_{0m} = 1$, and $\widetilde{\eta}_\Delta = (\eta_\Delta - \delta_0)/(\beta_{1m} - \beta_{0m})$. The first and second stages of our procedure rely on the fact that we can rewrite the model as

$$\begin{aligned} D &= \mathbb{1}\{X'\zeta_0 + \widetilde{\eta}_\Delta > 0\} \\ Y_k &= X'\beta_k + \varepsilon_k, \quad k \in \{0, 1\}, \end{aligned} \tag{3.2}$$

[13] We suppose that the constant is not included in $X$, so that $\varepsilon_0$ and $\varepsilon_1$ do not necessarily have mean zero.

where $Y_k$ is observed when $D = k$, $\widetilde{\eta}_\Delta$ is independent of $X$ and $E(\varepsilon_k|D = k, X)$ only depends on $X'\zeta_0$.[14] Besides, by Assumption 3.1, $X_1$ (resp. $X_2$) affects selection since $\zeta_{01} \neq 0$ (resp. $\zeta_{02} \neq 0$) but not directly the outcome $Y_0$ (resp. $Y_1$). Hence, Equations (3.2) correspond to Newey (2009)'s selection model and we follow his approach here. First, we estimate $\zeta_0$ by a single index estimator $\widehat{\zeta}$, for which we suppose Assumption 3.4 to be satisfied. This is the case of many semiparametric estimators, such as the one of Klein & Spady (1993) or Ichimura (1993). Secondly, we estimate $\beta_0$ and $\beta_1$ by series estimator, and suppose that it satisfies Assumption 3.5. Note that it is possible to prove this condition under more primitive assumptions (see Newey, 2008, p. S227).

**Assumption 3.4** *(Regularity of the first stage estimator) There exists $(\psi_i)_{1\leq i\leq n}$, i.i.d. random variables such that $E(\psi_1) = 0$, $E(\psi_1\psi_1')$ exists and is non singular and*

$$\widehat{\zeta} - \zeta_0 = \frac{1}{n}\sum_{i=1}^{n} \psi_i + o_P\left(\frac{1}{\sqrt{n}}\right).$$

**Assumption 3.5** *(Regularity of the second stage estimators) Let $k \in \{0, 1\}$, there exists $(\psi_{ki})_{1\leq i\leq n}$, i.i.d. random variables such that $E(\psi_{k1}) = 0$, $E(\psi_{k1}\psi_{k1}')$ exists and is non singular and*

$$\widehat{\beta_k} - \beta_k = \frac{1}{n}\sum_{i=1}^{n} \psi_{ki} + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Since $\gamma_0 = \beta_1 - \beta_0 - \zeta_0(\beta_{1m} - \beta_{0m})$, we then estimate $\gamma_0$ by

$$\widehat{\gamma} = \widehat{\beta_1} - \widehat{\beta_0} - \widehat{\zeta}(\widehat{\beta_{1m}} - \widehat{\beta_{0m}}).$$

It is easy to see that Assumptions 3.4 and 3.5 imply the root-n convergence and asymptotic normality of $\widehat{\gamma}$. The main difficulty actually lies in the estimation of $\delta_0$, which we now consider in a third stage.

Define, in a similar spirit as before (but with a slight abuse of notations), $T_i = X_i'\lambda_0$. The third stage of our procedure is based on Equation (2.7), which writes here as

$$\varepsilon_i - D_iT_i + \int_{t_0}^{T_i} q_0(u)du = \alpha_0 + D_i\delta_0 + \xi_i, \quad E(\xi_i|T_i) = 0, \tag{3.3}$$

where $q_0(u) = E(D|T = u)$. Let $\theta_0 = (\alpha_0, \delta_0)'$, $V_i = \varepsilon_i - D_iT_i + \int_{t_0}^{T_i} q_0(u)du$ and $W_i = (1, D_i)'$, so that $V_i = W_i'\theta_0 + \xi_i$. We estimate $\theta_0$ with an IV estimator which, for technical

[14] Indeed, $\varepsilon_k = \eta_k + \nu_k$ with $E(\nu_k|D = k, X) = 0$ by definition and $E(\eta_1|D = 1, X = x) = E(\eta_1|\widetilde{\eta}_\Delta > -x'\zeta_0)$ (and similarly for $k = 0$). Note that in general, $\varepsilon_k$ is not independent of $X$ because $\nu_k$ is not.

reasons, includes some trimming. We consider (unfeasible) instruments of the kind $S_i = \mathbb{1}\{X_i \in \mathcal{X}\}(1, h(T_i))'$, where $h(T_i) \in \mathbb{R}$ and $\mathcal{X}$ is a set strictly included in the support of $X$ and such that $\{x'\lambda_0, x \in \mathcal{X}\}$ is an interval.[15] Then $\theta_0 = E(S_1 W_1')^{-1} E(S_1 V_1)$, and we estimate it by

$$\widehat{\theta} = \left( \frac{1}{n} \sum_{i=1}^{n} \widehat{S}_i W_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \widehat{S}_i \widehat{V}_i \right),$$

where

$$\begin{aligned} \widehat{V}_i &= \widehat{\varepsilon}_i - D_i \widehat{T}_i + \int_{t_0}^{\widehat{T}_i} \widehat{q}(u, \widehat{\lambda}) du \\ \widehat{S}_i &= \mathbb{1}\{X_i \in \mathcal{X}\} \left(1, h\left(\widehat{T}_i\right)\right)' \end{aligned}$$

with $\widehat{\varepsilon}_i = Y_i - X_i'(D_i \widehat{\beta}_1 + (1 - D_i)\widehat{\beta}_0)$, $\widehat{\lambda} = \widehat{\beta}_0 - \widehat{\beta}_1 + \widehat{\gamma}$, $\widehat{T}_i = X_i' \widehat{\lambda}$ and

$$\widehat{q}(u, \lambda) = \frac{\sum_{i=1}^{n} D_i K\left(\frac{u - X_i'\lambda}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{u - X_i'\lambda}{h_n}\right)}. \tag{3.4}$$

where $K(.)$ is a kernel function and $h_n$ a smoothing parameter. The result on the third step estimator $\widehat{\theta}$ relies on the following conditions on $h(.)$ and $K(.)$.

**Assumption 3.6** *(Restrictions on the kernel) $K$ is nonnegative, zero outside a compact set, continuously twice differentiable on this compact set and satisfies $\int K(v)dv = 1$ and $\int vK(v)dv = 0$. Moreover, $K(.)$ and $K'(.)$ are zero on the boundary of this compact set.*

**Assumption 3.7** *(Regular instruments) $h(.)$ is twice differentiable and $|h''|$ is bounded.*

Assumption 3.6 is satisfied for instance by the quartic kernel $K(v) = (15/16)(1-v^2)^2 \mathbb{1}_{[-1,1]}(v)$. Assumption 3.7 is imposed to ensure that $\widehat{S}_i - S_i$ is small for large values of the sample size $n$ and behaves regularly.

**Theorem 3.1** *Suppose that $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and that Assumptions 2.1, 2.3, 2.7, 3.1-3.7 hold. Then*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0,\, E(S_1 W_1')^{-1} V(S_1 \xi_1 + \Omega_{11} + \Omega_{21}) E(W_1 S_1')^{-1}\right).$$

*where $\Omega_{11}$ is defined by Equation (7.8) in Appendix A and*

$$\Omega_{21} = S_1(1 - F_0(T_i))\mathbb{1}\{T_i \geq t_0\}(D_i - q_0(T_i))/f_0(T_i)$$

*where $F_0(.)$ and $f_0(.)$ denote respectively the cumulative distribution function and the density of $T_1$.*

[15] Hence, these instruments depend on $X_i$ and not on $T_i$ only. This is not an issue since actually, one can show that $E(\xi_i | X_i) = 0$.

Recalling that $\theta_0 = (\alpha_0, \delta_0)'$, this theorem guarantees that the final estimator of $\delta_0$ is root-n consistent and asymptotically normal. Its asymptotic variance depends on the three variables $\Omega_{11}$, $\Omega_{21}$ and $S_1\xi_1$. The first one corresponds to the contribution of the estimators of the first and second steps. The second one arises because of the nonparametric estimation of $q_0(.)$. The third one corresponds to the moment estimation of the linear instrumental model (3.3) in the last step.

## 4 Monte Carlo simulations

We investigate the finite-sample performance of the semiparametric estimators proposed in the preceding section. Namely, we simulate the following model:

$$\begin{aligned} Y_{0i} &= X_{2i}\beta_{02} + X_{3i}\beta_{03} + \eta_{0i} + \nu_{0i} \\ Y_{1i} &= X_{1i}\beta_{11} + X_{3i}\beta_{13} + \eta_{1i} + \nu_{1i} \\ D_i &= \mathbb{1}\{-\delta_0 + X_{1i}\beta_{11} - X_{2i}(\beta_{02} + \gamma_{02}) + X_{3i}(\beta_{13} - \beta_{03} - \gamma_{03}) + \eta_{1i} - \eta_{0i} > 0\}. \end{aligned}$$

The true values of the parameters are $\beta_{02} = \beta_{03} = 1$, $\beta_{11} = 2$, $\beta_{13} = 0.5$, $\gamma_{02} = 0.5$, $\gamma_{03} = -0.8$ and $\delta_0 = 0.8$, so that Assumption 3.1 is satisfied with $m = 1$. We simulate $X_{1i}$ and $X_{2i}$ independently and from a uniform distribution over $[0, 4]$, while $X_{3i}$ is a discrete regressor drawn from a Bernoulli distribution with parameter $p = 0.5$. We let $(\eta_{0i}, \eta_{1i})'$ be joint normal, with mean $\mu = (0, 0)'$ and variance $\Sigma$ such that $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = 0.5$. $(\nu_{0i}, \nu_{1i})'$ are drawn from a heteroskedastic normal distribution, with mean $\kappa = (0, 0)'$ and a conditional variance $\Omega(X)$ such that $\Omega_{11}(X) = \exp(X_2/5)$, $\Omega_{22}(X) = \exp(X_1/5)$ and $\Omega_{12}(X) = \Omega_{21}(X) = 0.5\sqrt{\Omega_{11}(X)\Omega_{22}(X)}$.

We implement the three-stages estimation procedure detailed in Section 3. More precisely, we estimate in the first stage $\zeta_0 = (\beta_1 - \beta_0 - \gamma_0)/\beta_{11}$ by Klein & Spady's (1993) semiparametric efficient estimator, with an adaptive gaussian kernel and local smoothing. In the second stage, we implement Newey's (2008) method in order to estimate separately $\beta_0$, $\beta_1$ and $\gamma_0$. The series estimator of the selection correction term was computed using the inverse Mills ratio transform (see Newey, 2008, equation (3.6)) and Legendre polynomials at order 6. Using Legendre polynomials instead of simple power series avoids numerical trouble due to multicollinearity. In the third stage, we finally implement our proposed estimator for $\delta_0$ with the quartic kernel suggested in Section 3 and a bandwidth $h_n = 0.5\sigma(\widehat{T})n^{-1/7}$, where $\sigma(\widehat{T})$ is the estimated standard deviation of $\widehat{T}$. We choose the function $h(x) = \Phi(a_0 + a_1 x)$ for the instruments, where $\Phi(.)$ denotes the normal cumulative distribution and $(a_0, a_1)$ are obtained by a probit of $D$ on $T$. Finally, no trimming was performed since it did not seem to improve the accuracy of the estimator in our setting.

The performance of the estimators for different sample sizes (namely $n = 500$, $n = 1,000$ and $n = 2,000$) are summarized in Table 1, which reports for each parameter the mean estimate, the standard deviation and the root mean squared error (RMSE).

| Sample size | Parameter | Mean | Standard error | RMSE |
|---|---|---|---|---|
| 500 | $\beta_{02}$ | 0.989 | 0.117 | 0.118 |
| | $\beta_{03}$ | 1.013 | 0.221 | 0.222 |
| | $\beta_{11}$ | 2.016 | 0.155 | 0.156 |
| | $\beta_{13}$ | 0.494 | 0.199 | 0.199 |
| | $\gamma_{02}$ | 0.456 | 0.181 | 0.186 |
| | $\gamma_{03}$ | -0.803 | 0.347 | 0.347 |
| | $\delta_0$ | 0.866 | 0.575 | 0.578 |
| 1,000 | $\beta_{02}$ | 0.992 | 0.085 | 0.085 |
| | $\beta_{03}$ | 0.996 | 0.157 | 0.157 |
| | $\beta_{11}$ | 2.006 | 0.105 | 0.105 |
| | $\beta_{13}$ | 0.506 | 0.146 | 0.146 |
| | $\gamma_{02}$ | 0.456 | 0.127 | 0.134 |
| | $\gamma_{03}$ | -0.781 | 0.24 | 0.240 |
| | $\delta_0$ | 0.865 | 0.393 | 0.399 |
| 2,000 | $\beta_{02}$ | 0.989 | 0.058 | 0.059 |
| | $\beta_{03}$ | 1.001 | 0.108 | 0.108 |
| | $\beta_{11}$ | 2.01 | 0.071 | 0.071 |
| | $\beta_{13}$ | 0.504 | 0.098 | 0.098 |
| | $\gamma_{02}$ | 0.471 | 0.088 | 0.093 |
| | $\gamma_{03}$ | -0.792 | 0.168 | 0.168 |
| | $\delta_0$ | 0.833 | 0.276 | 0.278 |

The results were obtained with 1,000 simulations for each sample size.

Table 1: Monte Carlo simulations

The results indicate that the semiparametric estimation procedure proposed in Section 3 performs fairly well in this context. In particular, although the last stage estimator of the non-pecuniary constant component $\widehat{\delta}$ is as expected less precise than the estimators $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\gamma}$, its finite sample performance still remains reasonable. In particular, although not negligible until $n = 1,000$, its bias seems to decrease quickly after.

# 5 Application to the decision to attend higher education

In this section, we apply our identification results and semiparametric method to estimate the relative importance of non-pecuniary components and monetary returns to education in the decision to attend higher education in France. We first briefly present in Subsection 5.1 the underlying theoretical schooling choice model on which we rely. Subsection 5.2 presents the data we use. Subsection 5.3 provides some details on the computation of the streams of earnings and on the implementation of our estimation method. Finally, Subsection 5.4 and 5.5 present the results and some robustness checks.

## 5.1 Decision to attend higher education and consumption value of schooling

We consider here a generalization of the Willis & Rosen's model (1979) which accounts for the consumption value of schooling.[16] After completing secondary education, individuals are assumed to decide either to enter directly the labor market with a high school degree ($k = 0$) or to attend higher education ($k = 1$). They are supposed to make their decision $D$ by comparing the expected discounted streams of future earnings related to each alternative. When entering the labor market, individuals receive a stream of log-earnings denoted by $Y_k^*$ for each alternative $k$, and such that

$$Y_k^* = \psi_k(X) + \eta_k + \nu_k,$$

where $\psi_k(.)$ is an unknown function of observed individual covariates $X$, $(\eta_0, \eta_1)$ are individual productivity terms which are supposed to be known by the individual at the time of her decision but unobserved by the econometrician and $(\nu_0, \nu_1)$ represent random shocks with means zero, which are unobserved by both the individual and the econometrician. The expected utility $\mathcal{U}_k$ of each schooling decision $k$ is supposed to be given by

$$\mathcal{U}_k = E(Y_k^*|X, \eta_k) + G_k(X),$$

[16]On a related ground, Carneiro et al. (2003) also estimate a generalization of the Willis and Rosen model accounting for non-pecuniary factors affecting the decision to attend college. Nevertheless, they rely on a completely different factor loadings framework, which is quite demanding in terms of identifying conditions. Apart from the existence of standard exclusion restrictions entering only the selection equation, they also hinge on the availability in the NLSY 79 (*National Longitudinal Survey of Youth 1979*) of five different cognitive ability measures in order to identify their factor model. Many datasets, including ours as well as e.g. the U.S. Current Population Survey, lack such measurements. See also Carneiro & Lee (2009) who estimate on the same dataset a semiparametric reduced-form model of college attendance decision based on an extension of Heckman & Vytlacil (2005).

where $G_k(X)$ denotes the *consumption value* associated with the schooling decision $k$.[17] After graduating from high school, the individual is supposed to make the decision which yields the highest expected utility. Thus, the selection equation corresponds exactly to Equation (2.1). Noteworthy, as opposed in particular to the U.S., tuition fees are very low in most of the French higher education institutions (on average around 200 euros per year over the period of interest). This suggests that $G_1 - G_0$, which would in principe also account for the direct costs of post-secondary schooling, can be interpreted in this context as a truly non-pecuniary component.

### 5.2 The data

We use French data from the *Generation 1992* and *Generation 1998* surveys in order to estimate the previous model of schooling choice.[18] The *Generation 1992* (resp. *Generation 1998*) survey consists of a large sample of 26,359 (resp. 22,021) individuals who left the French educational system in 1992 (resp. 1998) and were interviewed five years later. The two databases have the main advantage to contain information on both educational and labor market histories (over the first five years following the exit from the educational system). Furthermore, the surveys provide a set of individual covariates which are used as controls in our estimation procedure such as gender, place of birth, nationality, parents' profession, and place of residence. As most of the individual covariates are observed in both dataset, we exploit the pooled dataset hereafter.

Our subsample of interest is constituted of respondents having at least passed the national high school final examination. The labor market participation rate is fairly high for this subsample. For individuals leaving the schooling system in 1992, it is equal to 99.7% for males and 95.9% for females, while for those leaving education in 1998, it reaches 99.3% for males and 97.2% for females. Thus, we decide to keep both males and females in our final sample. Dropping individuals who only worked as temporary workers or who were out of the labor force during the observation length, for whom wages are not observed in the data, finally leaves us with a large sample of 24,474 individuals. Although not common in the semiparametric literature estimating generalized Roy models, working with many observations is especially well suited for the semiparametric estimation procedure to perform

[17]As opposed to the *investment value* of schooling, which corresponds in this case to the expected discounted stream of future log-earnings.

[18]Beffy et al. (2009) also rely on these data to estimate the influence of expected returns when choosing a college major.

well.[19] We report below some descriptive statistics for the subsample of interest, according to higher education attendance. 79.1% of our sample (with a slight increase over the period, respectively 77.4% for *Generation 1992* and 80.7% for *Generation 1998*) attended higher education after graduating from high school. Note that, in a same spirit as in Willis & Rosen (1979), we focus on higher education attendance and not graduation. Hence, higher education dropouts are included in the subsample of higher education attendees.

| | Higher education attendees | | High school level | |
|---|---|---|---|---|
| Variable | Mean | Std. dev. | Mean | Std. dev. |
| Initial monthly log wage (1992 French Francs) | 8.75 | 0.44 | 8.50 | 0.39 |
| Secondary schooling track | | | | |
| L | 0.15 | 0.36 | 0.04 | 0.19 |
| ES | 0.17 | 0.38 | 0.04 | 0.19 |
| S | 0.32 | 0.47 | 0.06 | 0.23 |
| Vocational | 0.04 | 0.20 | 0.66 | 0.47 |
| Technical | 0.32 | 0.46 | 0.21 | 0.41 |
| Born abroad | 0.03 | 0.16 | 0.02 | 0.15 |
| Father born abroad | 0.11 | 0.32 | 0.11 | 0.32 |
| Mother born abroad | 0.11 | 0.31 | 0.10 | 0.30 |
| Male | 0.47 | 0.5 | 0.49 | 0.50 |
| Father's profession | | | | |
| Farmer | 0.06 | 0.25 | 0.08 | 0.27 |
| Tradesman | 0.11 | 0.31 | 0.11 | 0.32 |
| Executive | 0.26 | 0.44 | 0.10 | 0.30 |
| Intermediate occupation | 0.12 | 0.32 | 0.09 | 0.29 |
| Blue-collar | 0.17 | 0.38 | 0.30 | 0.46 |
| White-collar | 0.21 | 0.41 | 0.25 | 0.44 |
| Age in $6^{th}$ grade | | | | |
| $\leq 10$ | 0.10 | 0.29 | 0.03 | 0.17 |
| 11 | 0.84 | 0.37 | 0.72 | 0.45 |
| $\geq 12$ | 0.07 | 0.25 | 0.25 | 0.43 |
| Paris region | 0.16 | 0.36 | 0.12 | 0.32 |
| Number of higher education years | 2.82 | 1.45 | / | / |
| Dropout rate | 0.16 | 0.37 | / | / |
| Number of observations | 19,365 | | 5,109 | |

Table 2: Descriptive statistics.

[19]Papers in this literature usually rely on the NLSY 79 (see Cunha & Heckman, 2007), resulting in samples of around 1,000 observations.

Functions $\psi_0(.), \psi_1(.)$ and $G(.)$ are assumed to depend on secondary schooling track, whether the student is born abroad (and similarly for her parents), year of entry into the labor market (1992 or 1998), gender, parental profession, age in $6^{th}$ grade (which is used as a proxy for ability)[20] and a dummy for living in Paris region. Aside from this common set of regressors, we also allow $\psi_0(.)$ (resp. $\psi_1(.)$) to depend on the average local log-earnings of high school (resp. higher education) graduates. These variables, which are computed from the French Labor Force Survey (1990-2000), are used as proxies for local labor market conditions (at the level of the French *departements*, which roughly correspond to U.S. counties).[21] We assume that the non-pecuniary component $G(.)$ does not depend on the average local log-earnings of high school graduates, following Carneiro et al. (2003). Indeed, while migration costs implies that labor market conditions in the places where individuals live while studying are likely to be correlated with the earnings perceived when entering the labor market, there is no obvious reason why these local labor market variables should enter the non-pecuniary factor $G(.)$. Importantly, because $G(.)$ is identified with this single exclusion restriction, we can check the validity of this instrumental strategy by letting $G(.)$ depend on log-earnings for higher education graduates and testing for the significance of the corresponding parameter.

### 5.3 Computation of the streams of earnings and estimation method

For each alternative, the discounted streams of log-earnings are set equal to

$$Y_k^* = \sum_{t=t_{0,k}}^{t_{0,k}+A} \tau^t y_{k,t},$$

where $y_{k,t}$ denotes the flow of log-earnings received during year $t$, $\tau$ denotes the annual discount factor and $A$ is the duration of active life. We account for the opportunity costs incurred when entering higher education by allowing the year of entry into the labor market ($t_{0,k}$) to vary according to the schooling choice. For a given period $t$, the earnings variable $y_{k,t}$ is either set equal to the log-wage $w_t$ earned during this period if the individual is employed at that time, or to the unemployment log-benefits $b_t$ if the latter is unemployed. We set the replacement rate equal to 0.7 as often done in the literature.

[20] Note the rationale behind using this variable as a proxy for ability lies in the fact that most of its variation stems from grade retention, which is especially common in France and mainly based on schooling performances.

[21] More precisely, these variables were constructed by taking the mean of local log-earnings over a 5-years time span centered respectively in 1992 or in 1998.

As already mentioned, we do not observe incomes at all periods in our data, so that we cannot compute $Y^* = DY_1^* + (1-D)Y_0^*$. Still, we can recover an expectation of this stream of income under additional assumptions on incomes dynamics. We suppose here that

$$y_{k,t} = \rho_k \mathbb{1}\{t \leq B\} + y_{k,t-1} + \nu_{k,t},$$

where $\rho_k$ denotes the alternative $k$-specific return to experience and $\nu_{k,t}$ is a degree $k$-specific unobserved individual productivity term which is assumed to be independently and identically distributed over time, with mean zero. We introduce the dummy $\mathbb{1}\{t \leq B\}$ to account for non significant marginal returns to experience after $B$ years of work (see, e.g., Kuruscu, 2006, for similar assumption on wage growth). We also suppose that $\nu_{k,t}$ is independent of $D$, so that $\rho_k$ is simply identified by $\rho_k = E(y_{k,t} - y_{k,t-1}|D = k)$, for $t \leq B$. Then, we can compute when $D = k$ the following predicted stream of income:

$$\begin{aligned} Y_k &= \widetilde{\tau}_k y_{D,t_{0,D}} + \rho_k C_k \\ &= E(Y_k^*|X, \eta_0, \eta_1, \nu_{k,t_{0,k}}), \end{aligned}$$

where $\widetilde{\tau}_k = \tau^{t_{0,k}} \left(\frac{1-\tau^{A+1}}{1-\tau}\right)$ and $C_k = \tau^{t_{0,k}} \left(\frac{\tau}{(1-\tau)^2}\right) \left(1 - (B+1)\tau^B + B\tau^{B+1}\right)$. The last equality implies that $E(Y_k|X, \eta_0, \eta_1) = E(Y_k^*|X, \eta_0, \eta_1)$. In other terms, the model may be written in terms of $Y_k$ instead of $Y_k^*$, and our identification strategy applies with $Y = DY_1 + (1-D)Y_0$ instead of the unobserved variable $Y^*$.

In practice, we set $\tau = 0.95$, $A = 45$ years, $B = 25$ years, $\rho_0 = 0.025$ and $\rho_1 = 0.042$. These latter values for $\rho_0$ and $\rho_1$ were obtained by regressing $y_{k,t_{0,k}+T_k} - y_{k,t_{0,k}}$ on the number of years $T_k$ for which the income is observed, on the subsample satisfying $D = k$. Alternative specifications on some of these parameters are considered in Subsection 5.5.

We estimate the model relying on the three-stages semiparametric procedure detailed in Section 3. More precisely, we use for the first step a mixture of probit (see e.g. Coppejans, 2001) with $K_1 = 3$ mixture components.[22] The second step is performed with Newey (2009)'s series estimator, with $K_2 = 9$ approximating terms. $\delta_0$ is finally estimated with the same specifications as in the Monte Carlo simulations.[23]

We also set estimate the distribution of the *ex ante* treatment effects $\Delta$, namely $F_\Delta(u) = E[F_{\eta_\Delta}(u + X'(\beta_0 - \beta_1))]$. For that purpose, we use the fact that, by (3.1),

$$P(D = 0|X) = F_{\eta_\Delta}\left(\delta_0 + X'\lambda_0\right).$$

[22]We did not rely on Klein & Spady (1993)'s estimator as we did in the Monte Carlo simulations since it becomes computationally cumbersome as the number of covariates increases.

[23]We estimated the model with several different values for the tuning parameters $K_1$, $K_2$ and the bandwidth $h$ used in the estimation of $q_0$ in the third step. Our final results are robust to these specifications.

Therefore, we can obtain an estimator $\widehat{F}_{\eta_\Delta}(.)$ on $[\widehat{\underline{M}}, \widehat{\overline{M}}]$, the estimated support of $\delta_0 + X'\lambda_0$, by regressing nonparametrically $1-D$ on the index $\widehat{\delta}+X'\widehat{\lambda}$. On $[\widehat{\overline{M}}, +\infty)$ (resp. $(-\infty, \widehat{\underline{M}}]$), we simply set estimate $F_{\eta_\Delta}(.)$ by $[\widehat{\overline{P}}, 1]$ (resp. $[0, \widehat{\underline{P}}]$), where $\widehat{\overline{P}}$ (resp. $\widehat{\underline{P}}$) is the supremum (resp. infimum) of $\widehat{F}_{\eta_\Delta}(.)$. Finally, we estimate $\underline{F_\Delta}(u)$ and $\overline{F_\Delta}(u)$ with the empirical analogs of (2.8) and (2.9). Bounds on the distribution of the *ex ante* treatment effects for the treated are estimated similarly, relying on (2.10). In practice, we consider a kernel estimator of $F_{\eta_\Delta}$ with a gaussian kernel, and a bandwidth $h_n = 1.6\sigma(\widehat{T})n^{-1/5}$.

### 5.4 Results

Table 3 below reports the parameter estimates relative to the non-pecuniary component $G(.)$, while the estimates of $(\zeta, \beta_0, \beta_1)$ are deferred to Appendix C (Table 6). Overall, the results for $\beta_0$ and $\beta_1$ display a similar pattern. In particular, the local average income variables on which we rely to identify the non-pecuniary factors have a strong positive effect, significant at the 1% level, on earnings. Similarly, individuals entering the labor market in 1998 (relative to 1992) have very significantly higher earnings, reflecting the business cycle. As expected, males also earn significantly more for both levels of qualification. However, some characteristics only affect the earnings of high school graduates or higher education attendees. This is in particular the case of vocational secondary schooling tracks (resp. majors in humanities) relative to technical tracks, which are positively (resp. negatively) related to earnings for high school graduates.[24] Conversely, parental profession affects more significantly the earnings of higher education attendees, with negative signs associated with blue collar professions for the father as well as with inactive, deceased or unemployed mother or father.

Several patterns emerge from the estimates of $G(.)$ displayed below. First, the results suggest that individuals attending a general secondary schooling track, relative to a technical track, value positively higher education attendance, with the related coefficients being significant at the 1% level.[25] Conversely, those getting a high school degree from a vocational major have a much lower probability to attend higher education, with a parameter being nevertheless only significant at the 10% level. This pattern is consistent with the fact that the courses which are given in vocational secondary schooling tracks and, to a lesser extent, in technical tracks, are much more oriented towards the labor market than they

[24] Aside from the main vocational track effect, the earnings of those graduating from a vocational secondary major are significantly less affected by the business cycle.

[25] Recall that $G(.) = G_0(.) - G_1(.)$, so that a negative sign for a given coefficient of $G(.)$ implies a positive valuation of higher education compared to high school graduation.

are in general tracks. The positive effect of entering the labor market in 1998 may actually reflect the enlargement of access to higher education which took place in France during the nineties. Individuals living in Paris region also have a higher probability to attend higher education through these non-pecuniary factors, reflecting similarly a supply-side effect.[26] Parental profession, in particular that of the father, has also a significant influence on the non-pecuniary determinants of the decision to attend higher education. For instance, for a given *ex ante* return to higher education, individuals whose father is employed, relative to a white-collar position, as an executive, a tradesman or in an intermediate occupation have a higher propensity to enroll in higher education. This pattern suggests that part of the intergenerational transmission of human capital acts through non-pecuniary factors affecting the higher education attendance decision. Interestingly also, for a given level of expected monetary returns, males have a significantly higher probability of attending higher education, with parameter significant at the 1% level. This may be seen as reflecting higher educational aspirations for males than for females, transiting in particular through differential parental attitudes towards boys and girls. Age in 6th grade, which is used as a proxy for schooling ability, also affects the attendance decision through non-pecuniary factors. Relative to those who were on time, individuals who were less than 10 (resp. more than 12) when entering high school have a significantly higher (resp. lower) probability to get some post-secondary education. These results may stem from a positive correlation between schooling ability and taste (or motivation) for schooling. Finally, the positive effects on higher education attendance of entering the labor market in 1998 and of living in the Paris region are significantly diminished (at the 10% level only) for the individuals graduating from a vocational high school. This result stresses once more the important explanatory power of the secondary schooling track. Importantly, the coefficient related to the local average income of higher education graduates is small and not significant at standard levels. This suggests that it is reasonable to exclude from the non-pecuniary factors the local average income of high school graduates, strengthening our confidence in the validity of our identification strategy.

[26]The greater Paris area is indeed characterized by a particularly important density of post-secondary institutions, covering a wide range of fields.

| Variable | Baseline specification |
|---|---|
| Constant ($\delta_0$) | -0.001 (0.127) |
| Local average income | |
| Higher education graduates | -0.012 (0.007) |
| High school graduates | 0 |
| Secondary schooling track | |
| L | -0.123*** (0.044) |
| ES | -0.15*** (0.047) |
| S | -0.148*** (0.045) |
| Vocational | 0.244* (0.147) |
| Technical | *Ref.* |
| Born abroad | -0.033* (0.018) |
| Father born abroad | -0.001 (0.011) |
| Mother born abroad | -0.01 (0.015) |
| Entering the labor market in 1998 (relative to 1992) | -0.09** (0.036) |
| Male | -0.051*** (0.01) |
| Father's profession | |
| Farmer | -0.013 (0.013) |
| Tradesman | -0.024** (0.011) |
| Executive | -0.056** (0.024) |
| Intermediate occupation | -0.036*** (0.013) |
| Blue collar | 0.001 (0.008) |
| Other | -0.018* (0.011) |
| White collar | *Ref.* |
| Mother's profession | |
| Farmer | 0.05 (0.033) |
| Tradesman | 0.001 (0.011) |
| Executive | -0.019* (0.011) |
| Intermediate occupation | -0.017 (0.011) |
| Blue collar | 0.012 (0.008) |
| Other | -0.014* (0.007) |
| White collar | *Ref.* |
| Age in 6th grade | |
| $\leq 10$ | -0.042* (0.022) |
| 11 | *Ref.* |
| $\geq 12$ | 0.053** (0.026) |
| Paris region | -0.033** (0.013) |
| Vocational $\times$ ... | |
| Entering the labor market in 1998 | 0.03* (0.018) |
| Male | 0.002 (0.014) |
| Paris region | 0.049* (0.027) |

Standard errors, presented in parentheses, were computed by bootstrap with 1,000 bootstrap sample replicates. Significativity levels: *** (1%), ** (5%) and * (10%).

Table 3: Determinants of non-pecuniary factors: parameter estimates

The estimated distributions of the *ex ante* returns to higher education are displayed in Figure 1, respectively for the whole sample and for the subsample of higher education attendees. The streams were divided by 1,000 for scaling reasons, so that these returns must be compared to values which range from 0.3 to 2. A first striking point is that both distributions are point identified for most values. Differences between the upper and lower bounds appear only for $u \geq 0.36$, and still for these values the identifying interval remains small until $u \simeq 0.65$. Second, a lower bound $\underline{E}$ on $E(Y_1 - Y_0)$ can be estimated, using the upper bound of the distribution. We obtain $\underline{E} \simeq 0.07$, which is quite large since it corresponds roughly to one standard deviation of $Y$. Third, the heterogeneity on these returns is also large. If we consider that the support of the distribution is $[-0.6, 0.7]$, this yields a range on the *ex ante* returns $E(Y_1 - Y_0|X, \eta_0, \eta_1)$ which is equivalent to the one of $Y$. This substantial *ex ante* dispersion of the returns to higher education is in line with the conclusion of Cunha & Heckman (2007, p. 887) on U.S. data.

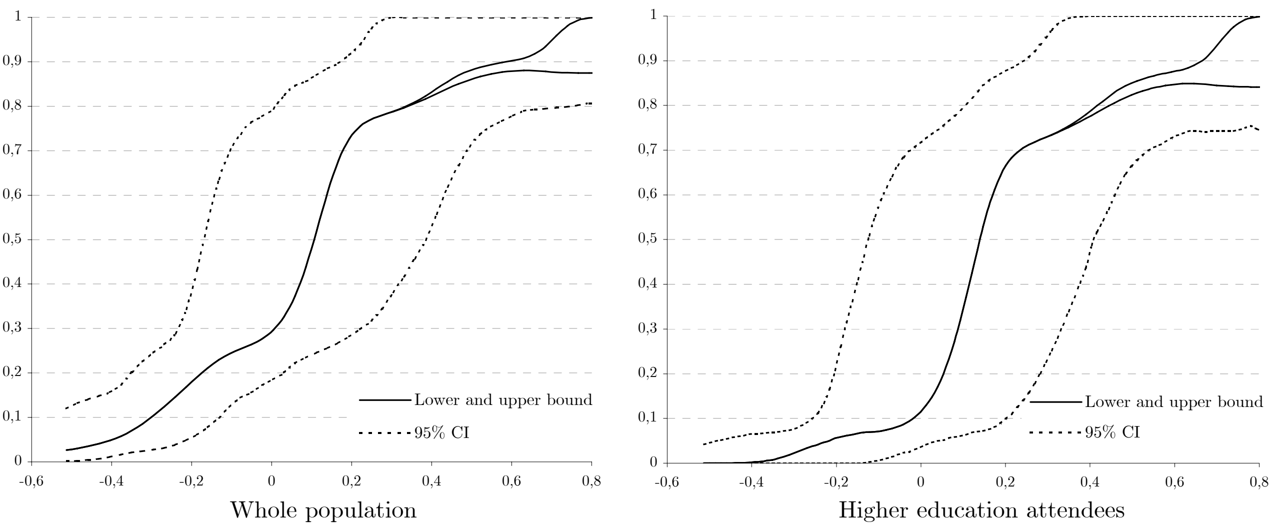


Figure 1: Distribution of the *ex ante* returns to higher education.

As expected, the distribution of the *ex ante* returns is shifted towards the right for the subsample of higher education attendees, with a close to 10% probability of having a negative *ex ante* return, versus 30% for the whole sample. Hence, about 10% of the individuals attending higher education choose to do so despite a negative *ex ante* return to higher education, stressing the important role played by non-pecuniary factors in this schooling decision. In a same spirit, denoting by $\widehat{F}_{\Delta}(.)$ the estimate of the cdf of the *ex ante* returns, we observe that $1 - \widehat{F}_{\Delta}(0) \simeq 70.6\%$. Taking the difference with the predicted access rate (82.1%) shows that the probability of attending higher education would fall by

11.5 percentage points if non-pecuniary factors did not exist. For comparison purposes, this decrease in higher education attendance rate is much larger than for instance the one associated with a 10% permanent decrease in labor market earnings of higher education attendees, namely 1.74 points only.

Several other results highlight the influence of non-pecuniary factors, relative to *ex ante* monetary returns, in the decision to attend higher education. First, as shown in Table 4 reporting the quartiles of the distribution of *ex ante* returns and non-pecuniary factors, the median non-pecuniary component (-0.263) is, in absolute terms, quantitatively much larger than the median *ex ante* return to higher education (0.106). Interestingly, the fact that the third quartile of the non-pecuniary component is negative suggests, in line with Carneiro et al. (2003), that there is for most of the individuals what could be referred to as a *psychic gain* of attending higher education.[27] Aside from their large median magnitude, non-pecuniary factors also have a fairly large dispersion, with an interquartile range equal to 0.201 which is nevertheless smaller than the interquartile range for *ex ante* returns (0.305).

| Quartile | *Ex ante* return | Non-pecuniary factors |
|---|---|---|
| 25% | -0.088 | -0.350 |
| 50% | 0.106 | -0.263 |
| 75% | 0.217 | -0.149 |

Table 4: Quartiles of *ex ante* returns and non-pecuniary factors.

Finally, Table 5 below reports the predicted probabilities of higher education attendance which are obtained for fixed values of the non-pecuniary factors corresponding respectively to the first and the last deciles of its sample distribution. These predicted attendance rates suggest once more that non-pecuniary factors matter much when deciding whether to attend higher education. Indeed, the predicted attendance rate falls steeply, by about 40 points, when making G vary from its first to its last decile. Overall, these results suggest that the variation across individuals in non-pecuniary factors accounts for a very substantial part of the observed decisions to attend higher education.

[27] Actually, it follows from the estimates of the non-pecuniary component that 83% of the individuals in the sample have a psychic gain of attending higher education.

| Decile of G | Predicted attendance rate |
|---|---|
| 10%, $G = -0.404$ | 0.952 |
| 90%, $G = 0.081$ | 0.573 |

Table 5: Predicted higher education attendance rates prevailing for different values of G.

### 5.5 Robustness checks

We address in the following three potential concerns about our results. The first one is the validity of our exclusion restriction. The second one is whether the generalized Roy model is a correct specification for the selection equation. The third one corresponds to the way we compute the stream of earnings.

#### *5.5.1 Validity of the instrumental strategy*

The non significance of the local average income of higher education graduates in the non-pecuniary component supports our exclusion restriction, but is still not a definitive proof of its validity. A reason why it might still not hold is that the decision to attend higher education could depend on local social norms in terms of educational attainment.[28] If places where earnings are higher were also those where the social gratification related to educational achievement is also higher, then the local labor market variables should not be excluded from $G(.)$. In order to cope with this potential concern, we include in the non-pecuniary component the local rate of higher education graduates relative to those with a secondary educational level or more. This rate, which is used to control for differences across *departements* in these social norms, is computed at the *departement* level from the French Census 1982 and 1990. The resulting estimates of $\gamma$ (see Panel 1, Table 7) are very similar to previously. Once more, the local average income of high school graduates does not affect the non-pecuniary factors component. Gender, father's and mother's profession and year of entry into the labor market remain the main determinants of this non-pecuniary component. The distribution of the *ex ante* returns to education is also very similar to previously (see Figure 2) and remains within the confidence intervals of that of the baseline

[28]Social norms may in particular act through social interactions on schooling choices. See in particular Cipollone & Rosolia (2007) and Lalive & Cattaneo (2009) for recent empirical evidence on this issue.

specification.[29]

#### 5.5.2 A specification test

The generalized Roy model considered here imposes a particular structure on the selection equation, which leads to Equation (3.3). As this equation is clearly overidentified, testing this structure is possible. To do so, let us remark that that if the model is true, then the regression between $V = \varepsilon - DT + \int_{t_0}^{T} q_0(u)du$ and $E(D|T)$ is linear. We thus consider a test of such a linear relationship against a nonparametric alternative. We implement the simple differencing test suggested by Yatchew (1998, p. 701) on $\widehat{V}$ and a kernel estimator of $E(D|T)$. We obtain a p-value of 2.24%, so that we do not reject the linear specification at the 1% level. Hence, assuming this selection rule seems to be reasonable given our data.[30]

#### 5.5.3 Alternative computations of the streams of earning

Finally, we also investigate the sensitivity of our results to the way the streams of earnings are computed. We reestimate the model with $\tau = 0.97$ instead of $\tau = 0.95$ (as, e.g., Carneiro et al., 2003), and $B = 30$ instead of $B = 25$. Results are displayed respectively in Panel 2 and 3 of Table 7. Once more, non-pecuniary components estimates are robust to this change. Standard errors, and thus the significance of some parameters, are slightly more affected by the specification choice. For instance, the local income of higher education graduates becomes significant at 10% (p-value=9%) when $B = 30$, as a result of a decrease of the standard error, the point estimates remaining stable throughout the different specifications. We also estimate the distribution of the *ex ante* returns to education with these alternative specifications (see Figure 3). Returns with $B = 30$ are nearly indistinguishable from the ones with $B = 25$. The distribution corresponding to $\tau = 0.97$ slightly dominates them, but remains within the confidence interval of the baseline specification. In a word, our results are overall robust to alternative computations of $Y$.[31]

[29] A reason why the estimates remain very stable is that the correlation between the local rate of higher education graduates and local average income variables is quite low (0.13 and 0.14 for high school and higher education graduates, respectively).

[30] Note also that the estimated cdf of the *ex ante* returns to higher education is increasing, which provides another check for the validity of our specification.

[31] We also estimate streams of earnings where people are aware of their own annual increase $\rho_i$ of log-earnings, instead of just anticipating an average increase. We estimate $\rho_i$ by OLS and compute the corresponding streams of earnings. The signs of $\gamma$ remain the same but no coefficient is significant anymore. This can be explained by 1) the importance of the errors on the estimated $\rho_i$ and 2) the fact that the sample we can use in this case comprises only 9,451 individuals.

# 6 Conclusion

This paper focuses on the effect of covariates on the potential outcomes and on the non-pecuniary component in a generalized Roy model. Our main contribution is to prove the nonparametric identification of the non-pecuniary component under the non standard assumption that at least one covariate affects the selection probability only through *ex ante* returns. In particular, local labor market conditions often appear to be natural candidates for this kind of exclusion restriction, which may be more convenient than the standard one in certain settings. We also contribute to the treatment effects literature by providing under this original instrumental strategy set identification results for the distribution of the treatment effects. We propose a three-stages semiparametric estimation procedure yielding root-n consistent and asymptotically normal estimators, the last stage allowing to estimate the non-pecuniary component from an instrumental linear model. Finally, relying on French data, we apply our method to quantify the relative importance of non-pecuniary components and expected returns to schooling in the decision to attend higher education. Consistently with the recent empirical evidence on this question, our results suggest that non-pecuniary factors have a major influence on the attendance decision.

Aside from applying our results to the analysis of, e.g., public versus private sector or migration decisions, another avenue for further research is the inference on the dependence between the sector-specific unobservable components $\eta_0$ and $\eta_1$. From an economic point of view, providing identification results on this dependence is especially worthwile since it conveys information about the relative importance of general vs. specific human capital. This dependence, which has received much attention in competing risks models (see, e.g., Peterson, 1976, van den Berg, 1997, Abbring & van den Berg, 2003), has been identified in generalized Roy models by imposing a factor model (see Carneiro et al., 2003). However, it would be interesting to conduct a more flexible analysis on this issue, without assuming that the outcomes depend on a low-dimensional set of factors. We leave this question for further research.

# 7 Appendix A: proofs

### Theorem 2.1

Recall that $\varepsilon_k = \eta_k + \nu_k$ for $k \in \{0,1\}$. Because $E(\nu_k|X, \eta_0, \eta_1) = 0$, we have $E(\nu_k|X, D = k) = 0$. Thus, by Assumptions 2.1 and 2.3,

$$\begin{aligned} E(\varepsilon_1|D=1, X=x) &= \frac{E(\eta_1 D|X=x)}{P(D=1|X=x)} \\ &= \frac{E\left(\eta_1 \mathbb{1}\{\eta_\Delta \geq \psi_0(x) - \psi_1(x) + G(x)\}\right)}{P(D=1|X=x)} \end{aligned} \quad (7.1)$$

Now let us show that almost surely,

$$\eta_\Delta \geq \psi_0(x) - \psi_1(x) + G(x) \iff S_{\eta_\Delta}(\eta_\Delta) \leq P(D=1|X=x) \quad (7.2)$$

where $S_{\eta_\Delta}$ denotes the survival function of $\eta_\Delta$. The first implication is obvious since $S_{\eta_\Delta}$ is decreasing. Now suppose that $S_{\eta_\Delta}(\eta_\Delta) \leq P(D=1|X=x)$. Then $\eta_\Delta \geq \inf A_x$ where $A_x = \{u/S_{\eta_\Delta}(u) = P(D=1|X=x)\}$. Now, for all interval $I \subset A_x$, $P(\eta_\Delta \in I) = 0$ by definition of $A_x$. Hence, because $\psi_0(x) - \psi_1(x) + G(x) \in A_x$, almost surely,

$$\eta_\Delta \geq \inf A_x \Rightarrow \eta_\Delta \geq \psi_0(x) - \psi_1(x) + G(x).$$

Hence, (7.2) holds. Then, by (7.1),

$$E(\varepsilon_1|D=1, X=x) = \frac{E(\eta_1 \mathbb{1}\{S_{\eta_\Delta}(\eta_\Delta) \leq P(D=1|X=x)\})}{P(D=1|X=x)}$$

In other terms, there exists a measurable function $h$ such that $E(\varepsilon_1|D=1, X) = h(P(D=1|X))$. Now, by Assumption 2.4,

$$E(Y|D=1, X) = \psi_1(X_1, X_c) + h(P(D=1|X)).$$

Suppose that there exists $\widetilde{\psi_1}$ and $\widetilde{h}$ such that

$$E(Y|D=1, X) = \widetilde{\psi_1}(X_1, X_c) + \widetilde{h}(P(D=1|X)).$$

Then

$$(\widetilde{\psi_1} - \psi_1)(X_1, X_c) + (\widetilde{h} - h)(P(D=1|X)) = 0$$

By the measurably separation condition, this implies that $\widetilde{\psi_1}$ and $\psi_1$ are almost surely equal up to a constant. This constant is identified by Assumption 2.2. Thus, $\psi_1$ is identified. $\psi_0$ can be recovered by the same argument.

**Theorem 2.2**

The proof relies on Theorem 2.1 of d'Haultfoeuille & Maurel (2009). Assumptions 1 and 2 of d'Haultfoeuille & Maurel (2009) are satisfied by conditions (i) and (ii) of Assumption 2.5. All we have to check is that Assumption 3 also holds. For that purpose, remark that for $k \in \{0, 1\}$,

$$\begin{aligned} P(D = k|X = x, Y_k = y) &= P(D = k|X = x, \varepsilon_k = y - \psi_k(x)) \\ &= P(\eta_k - \eta_{1-k} > \psi_{1-k}(x) - \psi_k(x) + G(x)|\eta_k + \nu_k = y - \psi_k(x)) \end{aligned}$$

Thus, by Condition (iii) of Assumption 2.5,

$$\lim_{y\to\infty} P(D = k|X = x, Y_k = y) = 1, \text{ for all } x.$$

In other words, Assumption 3 of d'Haultfoeuille & Maurel (2009) holds, so that the result follows.

**Theorem 3.1**

Before proving the results, let us introduce some notations. Let $U_i$ denote all the data corresponding to individual $i$, let $f(., \lambda)$ denote the density of $X'\lambda$, $q(u, \lambda) = E(D|X'\lambda = u)$, $r(., \lambda) = q(., \lambda) \times f(., \lambda)$ and define $f_0(u) = f(u, \lambda_0)$, $q_0(u) = q(u, \lambda_0)$ and $r_0(u) = q_0(u)f_0(u)$. Consider the kernel estimators

$$\widehat{f}(u, \lambda) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{u - X_i'\lambda}{h_n}\right)$$

and $\widehat{r}(., \lambda) = \widehat{q}(., \lambda) \times \widehat{f}(., \lambda)$, where $\widehat{q}(., \lambda)$ is defined by Equation (3.4). Let us also define $S_i(\lambda) = \mathbb{1}\{X_i \in \mathcal{X}\}(1, h(X_i'\lambda))'$ and, for any $\mu = (r(.), f(.), \lambda, \widetilde{\beta}_0, \widetilde{\beta}_1)$,

$$V_i(\mu) = Y_i - X_i'(D_i\widetilde{\beta}_1 + (1 - D_i)\widetilde{\beta}_0) - D_iX_i'\lambda + \int_{t_0}^{X_i'\lambda} \frac{r(u)}{f(u)}du.$$

Thus, $\widehat{V}_i = V_i(\widehat{\mu})$ and $V_i = V_i(\mu_0)$, with $\widehat{\mu} = (\widehat{r}(., \widehat{\lambda}), \widehat{f}(., \widehat{\lambda}), \widehat{\lambda}, \widehat{\beta}_0, \widehat{\beta}_1)$ and $\mu_0 = (r_0, f_0, \lambda_0, \beta_0, \beta_1)$. Eventually, let $g(U_i, \theta, \mu) = S_i(\lambda)(V_i(\mu) - W_i'\theta)$ and $g(U_i, \mu) = g(U_i, \theta_0, \mu)$. Then $E[g(U_1, \mu_0)] = 0$ and

$$\sum_{i=1}^{n} g(U_i, \widehat{\theta}, \widehat{\mu}) = 0.$$

Thus, $\widehat{\theta}$ is a two step GMM estimator with a nonparametric first step estimator, and we follow Newey & McFadden (1994)'s outline for establishing asymptotic normality. Some differences arise however because of the estimation of $\lambda$ in the nonparametric estimator of $q_0$. The proof of the theorem proceeds in three steps.

Step 1. We first show that $\mu \mapsto \sum_{i=1}^n g(U_i, \mu)$ can be linearized in a convenient way. Let

$$\begin{aligned} G(U_i, \mu) &= \xi_i \frac{\partial S_i}{\partial \lambda}(\lambda_0)'\lambda + S_i(\lambda_0)\Bigg[ - X_i'(D_i\widetilde{\beta}_1 + (1-D_i)\widetilde{\beta}_0) - D_i X_i'\lambda \\ &\quad + q_0(T_i)X_i'\lambda + \int_{t_0}^{T_i} \frac{\partial q}{\partial \lambda}(u, \lambda_0)'\lambda + \frac{1}{f_0(u)}\left(r(u) - q_0(u) f(u)\right) du \Bigg] \end{aligned}$$

Note that $\partial q/\partial\lambda(., \lambda_0)$ exists under Assumptions 2.3 and 3.2, by Lemma 8.1. Let us also define $\widetilde{\mu} = (\widetilde{r}, \widetilde{f}, \widehat{\lambda}, \widehat{\beta}_0, \widehat{\beta}_1)$ where $\widetilde{r} = \widehat{r}(., \lambda_0)$ and $\widetilde{f} = \widehat{f}(., \lambda_0)$. We shall prove that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[g(U_i, \widehat{\mu}) - g(U_i, \mu_0) - G(U_i, \widetilde{\mu} - \mu_0)\right] = o_P(1). \tag{7.3}$$

For that purpose, we use the decomposition

$$g(U_i, \widehat{\mu}) - g(U_i, \mu_0) - G(U_i, \widetilde{\mu} - \mu_0) = R_{1i} + R_{2i} + R_{3i} + R_{4i} + R_{5i}$$

where

$$\begin{aligned} R_{1i} &= \left(0, \xi_i \mathbb{1}\{X_i \in \mathcal{X}\}\left(h(\widehat{T}_i) - h(T_i) - h'(T_i)(\widehat{T}_i - T_i)\right)\right)' \\ R_{2i} &= S_i(\lambda_0)\left[\int_{T_i}^{\widehat{T}_i} \widehat{q}(u, \widehat{\lambda})du - q_0(T_i)(\widehat{T}_i - T_i)\right] \\ R_{3i} &= S_i(\lambda_0)\int_{t_0}^{T_i} \widehat{q}(u, \widehat{\lambda}) - \widetilde{q}(u) - \frac{\partial q}{\partial \lambda}(u, \lambda_0)(\widehat{\lambda} - \lambda_0)du \\ R_{4i} &= S_i(\lambda_0)\int_{t_0}^{T_i} \widetilde{q}(u) - q_0(u) - \frac{1}{f_0(u)}\left(\widetilde{r}(u) - r_0(u) - q_0(u)(\widetilde{f}(u) - f_0(u))\right) du \\ R_{5i} &= (V_i(\widehat{\mu}) - V_i(\mu_0))\left(S_i(\widehat{\lambda}) - S_i(\lambda_0)\right) \end{aligned}$$

where $\widetilde{q} = \widetilde{r}/\widetilde{f}$. We now check that for all $k \in \{1, ..., 5\}$, $1/\sqrt{n}\sum_{i=1}^n R_{ki} = o_P(1)$.

$-$ $R_{1i}$: by Assumption 3.2 and the Cauchy-Schwartz inequality, there exists $C_0 > 0$ such that $|\widehat{T}_i - T_i| \leq C_0||\widehat{\lambda} - \lambda_0||$. Thus, by Assumption 3.7,

$$\begin{aligned} \sqrt{n} \max_{i=1,\dots,n} \left|h(X_i\widehat{\lambda}) - h(T_i) - h'(T_i)(\widehat{T}_i - T_i)\right| &\leq \sqrt{n}M \max_{i=1,\dots,n} |\widehat{T}_i - T_i|^2 \\ &\leq MC_0^2\sqrt{n}||\widehat{\lambda} - \lambda_0||^2 \\ &= o_P(1), \end{aligned}$$

where $M$ denotes an upper bound of $|h''|$. Besides, $\sum_{i=1}^n |\xi_i|/n = O_P(1)$. Thus,

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{1i}\right\| = o_P(1).$$

$-\, R_{2i}$: Let $\mathcal{S}_0 = \{x'\lambda_0, x \in \mathcal{X}\}$. Because $\mathcal{X}$ is strictly included in the support of $X_1$, $\mathcal{S}_0 \subsetneq \mathcal{S}$. Besides, by definition of $S_i(\lambda_0)$, $S_i(\lambda_0) = S_i(\lambda_0)\mathbb{1}\{T_i \in \mathcal{S}_0\}$. Moreover, for all $i$ such that $\widehat{T}_i \in \mathcal{S}_0$, there exists, by the mean value theorem, $\widetilde{T}_i = tT_i + (1-t)\widehat{T}_i$, with $t \in [0,1]$, such that $\int_{T_i}^{\widehat{T}_i} q_0(u)du = q_0(\widetilde{T}_i)(\widehat{T}_i - T_i)$. Thus, when $\widehat{T}_i \in \mathcal{S}_0$,

$$
\begin{aligned}
\|R_{2i}\| &= \left\| S_i(\lambda_0)\mathbb{1}\{T_i \in \mathcal{S}_0\} \left\{ \int_{T_i}^{\widehat{T}_i} \left[\widehat{q}(u,\widehat{\lambda}) - q_0(u)\right] du + \int_{T_i}^{\widehat{T}_i} q_0(u) - q_0(T_i)(\widehat{T}_i - T_i) \right\} \right\| \\
&\leq C_1 \left|\widehat{T}_i - T_i\right| \left[ \sup_{u\in\mathcal{S}_0} \left|\widehat{q}(u,\widehat{\lambda}) - q_0(u)\right| + \max_{i:\widehat{T}_i\in\mathcal{S}} \left|q_0(\widetilde{T}_i) - q_0(T_i)\right| \right] \\
&\leq C_0C_1 \left\|\widehat{\lambda} - \lambda_0\right\| \left[ \sup_{u\in\mathcal{S}_0} \left|\widehat{q}(u,\widehat{\lambda}) - q_0(u)\right| + \max_{i:\widehat{T}_i\in\mathcal{S}} \left|q_0(\widetilde{T}_i) - q_0(T_i)\right| \right],
\end{aligned}
$$

where $C_1 > 0$ is a constant such that $\|S_i(\lambda_0)\| \leq C_1$. Besides, because $\widehat{q}(.,\widehat{\lambda})$ and $q_0(.)$ are bounded by 1, we have, when $\widehat{T}_i \notin \mathcal{S}_0$,

$$
\|R_{2i}\| \leq 2C_0C_1 \left\|\widehat{\lambda} - \lambda_0\right\| \mathbb{1}\{T_i \in \mathcal{S}_0\}.
$$

Hence,

$$
\begin{aligned}
\|R_{2i}\| \;\leq\; & C_0C_1 \left\|\widehat{\lambda} - \lambda_0\right\| \left[ \sup_{u\in\mathcal{S}_0} \left|\widehat{q}(u,\widehat{\lambda}) - q_0(u)\right| + \max_{i:\widehat{T}_i\in\mathcal{S}} \left|q_0(\widetilde{T}_i) - q_0(T_i)\right| \right. \\
& \left. +2\mathbb{1}\{T_i \in \mathcal{S}_0, \widehat{T}_i \notin \mathcal{S}_0\} \right]. \qquad (7.4)
\end{aligned}
$$

By Assumption 3.4 and 3.5, $\sqrt{n}\left\|\widehat{\lambda} - \lambda_0\right\| = O_P(1)$. Let us now consider the term into brackets in (7.4). By Lemma 8.2, $\sup_{u\in\mathcal{S}_0} |\widehat{q}(u,\widehat{\lambda}) - q_0(u)| = o_P(1)$. Let us prove that

$$
\max_{i:\widehat{T}_i\in\mathcal{S}} \left|q_0(\widetilde{T}_i) - q_0(T_i)\right| = o_P(1) \qquad (7.5)
$$

Fix $\varepsilon > 0$. Because $q_0(.)$ is continuous by Assumption 3.2 and $\mathcal{S}$ is compact, $q_0(.)$ is uniformly continuous on $\mathcal{S}$. Thus, there exists $\delta > 0$ such that for all $(u,v) \in \mathcal{S}^2$ satisfying $|u - v| \leq \delta$, we have $|q_0(u) - q_0(v)| \leq \varepsilon$. As a consequence,

$$
P\left(\max_{i:\widehat{T}_i\in\mathcal{S}} \left|q_0(\widetilde{T}_i) - q_0(T_i)\right| \leq \varepsilon\right) \geq P\left(\max_{i:\widehat{T}_i\in\mathcal{S}} \left|\widetilde{T}_i - T_i\right| \leq \delta\right).
$$

Because $|\widetilde{T}_i - T_i| \leq |\widehat{T}_i - T_i| \leq C_0\left\|\widehat{\lambda} - \lambda_0\right\|$, the right-hand side tends to one. This establishes (7.5). It remains to show that

$$
\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{T_i \in \mathcal{S}_0, \widehat{T}_i \notin \mathcal{S}_0\} = o_P(1). \qquad (7.6)
$$

For all $\delta > 0$, let $\mathcal{S}_\delta = \{s \in \mathcal{S}_0 / \exists s' \notin \mathcal{S}_0 / |s - s'| < \delta\}$. Fix $\varepsilon > 0$ and let $K > 0$ be such that $P(T_i \in \mathcal{S}_K) < \varepsilon/2$. For $n$ large enough, $P(C_0 \left\|\widehat{\lambda} - \lambda_0\right\| > K) < \varepsilon/2$. Because $|T_i - \widehat{T}_i| < C_0 \left\|\widehat{\lambda} - \lambda_0\right\|$, we have, for $n$ large enough,

$$\begin{aligned} P\left(T_i \in \mathcal{S}_0, \widehat{T}_i \notin \mathcal{S}_0\right) &\leq \frac{\varepsilon}{2} + P\left(T_i \in \mathcal{S}_0, \widehat{T}_i \notin \mathcal{S}_0, C_0 ||\widehat{\lambda} - \lambda_0|| \leq K\right) \\ &\leq \frac{\varepsilon}{2} + P(T_i \in \mathcal{S}_K) \\ &\leq \varepsilon. \end{aligned}$$

Because $\varepsilon$ was arbitrary, this proves that

$$E\left[\left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{T_i \in \mathcal{S}_0, \widehat{T}_i \notin \mathcal{S}_0\}\right|\right] \to 0.$$

This establishes (7.6) since convergence in $L^1$ implies convergence in probability. As a result, $\sum_{i=1}^n R_{2i}/\sqrt{n} = o_P(1)$.

$-$ $R_{3i}$: By the mean value theorem, there exists $\widetilde{\lambda}_u$ in the segment between $\lambda_0$ and $\widehat{\lambda}$ such that

$$\widehat{q}(u, \widehat{\lambda}) - \widetilde{q}(u) = \frac{\partial \widehat{q}}{\partial \lambda}(u, \widetilde{\lambda}_u)'(\widehat{\lambda} - \lambda_0).$$

Because $T_i$ is bounded, there exists $C_2$ such that $|T_i - t_0| < C_2$. Thus,

$$\begin{aligned} |R_{3i}| &= \|S_i(\lambda_0)\| \left| \left[\int_{t_0}^{T_i} \frac{\partial \widehat{q}}{\partial \lambda}(u, \widetilde{\lambda}_u) - \frac{\partial q}{\partial \lambda}(u, \lambda_0) du\right]' (\widehat{\lambda} - \lambda_0)\right| \mathbb{1}\{T_i \in \mathcal{S}_0\} \\ &\leq C_1 C_2 \left\|\widehat{\lambda} - \lambda_0\right\| \sup_{u \in \mathcal{S}_0} \left\|\frac{\partial \widehat{q}}{\partial \lambda}(u, \widetilde{\lambda}_u) - \frac{\partial q}{\partial \lambda}(u, \lambda_0)\right\|. \end{aligned}$$

The supremum tends to zero in probability by Lemma 8.2. As a result, $\sum_{i=1}^n R_{3i}/\sqrt{n} = o_P(1)$.

$-$ $R_{4i}$: following Newey & McFadden (1994, p. 2204), we have

$$\begin{aligned} |R_{4i}| &\leq C_1 \mathbb{1}\{T_i \in \mathcal{S}_0\} \int_{t_0}^{T_i} \frac{1}{\widetilde{f}(u) f_0(u)} \left[1 + |q_0(u)|\right] \left[|\widetilde{f}(u) - f_0(u)|^2 + |\widetilde{r}(u) - r_0(u)|^2\right] du \\ &\leq \frac{2 C_1 C_2}{\inf_{u \in \mathcal{S}_0} \widetilde{f}(u) \inf_{u \in \mathcal{S}_0} f_0(u)} \left[\left(\sup_{u \in \mathcal{S}_0} |\widetilde{f}(u) - f_0(u)|\right)^2 + \left(\sup_{u \in \mathcal{S}_0} |\widetilde{r}(u) - r_0(u)|\right)^2\right]. \end{aligned}$$

Assumption 3.2 implies that the density of $T_i$ is positive in the interior of $\mathcal{S}$. Thus, $\inf_{u \in \mathcal{S}_0} f_0(u) > 0$. By uniform consistency of $\widetilde{f}$ on $\mathcal{S}_0$ (see, e.g., Lemma 8.10 of Newey & McFadden, 1994) the ratio is a $O_P(1)$. Thus it suffices to show that $\sup_{u \in \mathcal{S}_0} |\widetilde{f}(u) - f_0(u)| = o_P(n^{-1/4})$ and similarly for $\widetilde{r}$. The result follows from Assumption 3.6, the rate condition on $h_n$ and Lemma 8.10 of Newey & McFadden (1994).

$- R_{5i}$: first, note that

$$\begin{aligned}
|V_i(\widehat{\mu}) - V_i(\mu_0)| \mathbb{1}\{X_i \in \mathcal{X}\} &= \Big| X_i'(D_i(\beta_1 - \widehat{\beta}_1) + (1 - D_i)(\beta_0 - \widehat{\beta}_0)) + D_i(T_i - \widehat{T}_i) \\
&\quad + \int_{T_i}^{\widehat{T}_i} \widehat{q}(u, \widehat{\lambda}) du + \int_{t_0}^{T_i} \left[ \widehat{q}(u, \widehat{\lambda}) - q_0(u) \right] du \Big| \mathbb{1}\{X_i \in \mathcal{X}\} \\
&\leq C_0 \left( \left\| \widehat{\beta}_1 - \beta_1 \right\| + \left\| \widehat{\beta}_0 - \beta_0 \right\| + 2 \left\| \widehat{\lambda} - \lambda_0 \right\| \right) \\
&\quad + C_2 \sup_{u \in \mathcal{S}_0} |\widehat{q}(u, \widehat{\lambda}) - q_0(u)|.
\end{aligned}$$

With probability approaching one, there exists a compact which contains $\widehat{T}_i$ and $T_i$ for all $i$. Thus, because $h'$ is continuous, there exists $C_3 > 0$ such that, with probability approaching one,

$$\left\| S_i(\widehat{\lambda}) - S_i(\lambda_0) \right\| \leq C_3 \left\| \widehat{\lambda} - \lambda_0 \right\|.$$

Thus, with probability approaching one,

$$\begin{aligned}
\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n R_{5i} \right| &\leq \left[ C_0 C_3 \sqrt{n} \left\| \widehat{\lambda} - \lambda_0 \right\| \right] \left[ \left\| \widehat{\beta}_1 - \beta_1 \right\| + \left\| \widehat{\beta}_1 - \beta_1 \right\| + 2 \left\| \widehat{\lambda} - \lambda_0 \right\| \right. \\
&\quad \left. + C_2 \sup_{u \in \mathcal{S}_0} |\widehat{q}(u, \widehat{\lambda}) - q_0(u)| \right].
\end{aligned}$$

By Assumptions 3.4 and 3.5, the first term into brackets in the right-hand side is a $O_P(1)$. By Lemma 8.2 and Assumptions 3.4 and 3.5, the second term is a $o_P(1)$. The result follows.

Step 2. Now, let us show that $1/\sqrt{n} \sum_{i=1}^n G(U_i, \widetilde{\mu} - \mu_0)$ can be linearized. Let $\kappa_0 = (\lambda_0, \beta_1, \beta_0)'$ and $\widehat{\kappa} = (\widehat{\lambda}, \widehat{\beta}_1, \widehat{\beta}_0)'$. We have

$$G(U_i, \widetilde{\mu} - \mu_0) = P_i' (\widehat{\kappa} - \kappa_0) + \widetilde{G}(U_i, \widetilde{r}, \widetilde{f}),$$

with $P_i = (P_{1i}, P_{2i}, P_{3i})'$ and

$$\begin{aligned}
P_{1i} &= (V_i(\mu_0) - W_i'\theta_0) \frac{\partial S_i}{\partial \lambda}(\lambda_0)' - S_i(\lambda_0) \left( D_i X_i' + q_0(T_i) X_i' + \int_{t_0}^{T_i} \frac{\partial q}{\partial \lambda'}(u, \lambda_0) du \right) \\
P_{2i} &= -D_i S_i(\lambda_0) X_i' \\
P_{3i} &= -(1 - D_i) S_i(\lambda_0) X_i' \\
\widetilde{G}(U_i, \widetilde{r}, \widetilde{f}) &= S_i(\lambda_0) \int_{t_0}^{T_i} (1/f_0(u))(\widetilde{r}(u) - q_0(u)\widetilde{f}(u)) du.
\end{aligned}$$

By the weak law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n P_i \xrightarrow{P} E[P_1].$$

Moreover, we have $\widehat{\lambda} = (\widehat{\beta}_{0m} - \widehat{\beta}_{1m})\widehat{\zeta}$. Thus, by Assumptions 3.4 and 3.5,

$$\widehat{\lambda} - \lambda_0 = \frac{1}{n}\sum_{i=1}^n \widetilde{\psi}_i + o_P\left(\frac{1}{\sqrt{n}}\right),$$

where, letting $\psi_{k1i}$ denote the first component of $\psi_{ki}$ for $k \in \{0, 1\}$,

$$\widetilde{\psi}_i = (\psi_{0mi} - \psi_{1mi})\zeta + (\beta_{0m} - \beta_{1m})\psi_i.$$

Hence,

$$\widehat{\kappa} - \kappa_0 = \frac{1}{n}\sum_{i=1}^n \left(\widetilde{\psi}_i, \psi_{1i}, \psi_{0i}\right)' + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Thus,

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n P_i\right)' (\widehat{\kappa} - \kappa_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Omega_{1i} + o_P(1), \tag{7.7}$$

where

$$\Omega_{1i} = E[P_1]' \left(\widetilde{\psi}_i, \psi_{1i}, \psi_{0i}\right)'. \tag{7.8}$$

Thus, it suffices to focus on the nonparametric part of $G$, $\widetilde{G}(U_i, \widetilde{r}, \widetilde{f})$. Now, $\widetilde{G}$ is nearly the linearized part of the consumer surplus example of Newey & McFadden (1994, p. 2204), except that $b$ is replaced by $T_i$. Thus, it suffices to modify slightly their proof (see Newey & McFadden, 1994, p. 2211) to satisfy Conditions (ii), (iii) and (iv) as well as the technical requirements of their Theorem 8.11. As a result, we get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{G}(U_i, r, f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Omega_{2i} + o_P(1), \tag{7.9}$$

where $\Omega_{2i} = S_i(\lambda_0)(1-F_0(T_i))\mathbb{1}\{T_i \geq t_0\}(D_i - q_0(T_i))/f_0(T_i)$, $F_0(.)$ denoting the cumulative distribution function of $T$. The result follows.

Step 3. Eventually, we establish the asymptotic normality of $\widehat{\theta}$. By (7.3), (7.7) and (7.9) and the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n g(U_i, \widehat{\mu}) \xrightarrow{d} \mathcal{N}\left(0,\, V(g(U_1, \mu_0) + \Omega_{11} + \Omega_{21})\right).$$

Thus, by definition of $\widehat{\theta}$ and $g(U_i, \theta, \mu)$,

$$\left[\frac{1}{n}\sum_{i=1}^n S_i(\widehat{\lambda})W_i'\right] \sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0,\, V(g(U_1, \mu_0) + \Omega_{11} + \Omega_{21})\right).$$

Now, recalling that by Assumption 3.7, $\left\|S_i(\widehat{\lambda}) - S_i(\lambda_0)\right\| \leq C_3 \left\|\widehat{\lambda} - \lambda_0\right\|$ for a given $C_3 > 0$. Thus,

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n} S_i(\widehat{\lambda})W_i' - E(S_1(\lambda_0)W_1')\right\| \quad &\leq \quad C_3\left(\frac{1}{n}\sum_{i=1}^{n}\|W_i\|\right)\left\|\widehat{\lambda} - \lambda_0\right\| \\ &\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} S_i(\lambda_0)W_i' - E(S_1(\lambda_0)W_1')\right\|.\end{aligned}$$

Thus, by the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} S_i(\widehat{\lambda})W_i' \xrightarrow{P} E(S_1(\lambda_0)W_1') = E(S_1W_1').$$

Eventually, by Slutski's lemma, and given that $g(U_1, \mu_0) = S_1\xi_1$,

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0,\ E(S_1W_1')^{-1}V(S_1\xi_1 + \Omega_{11} + \Omega_{21})E(W_1S_1')^{-1}\right).$$

This concludes the proof.

# 8 Appendix B: technical lemmas

**Lemma 8.1** *Suppose that Assumptions 2.3 and 3.2 hold. Then, for all $u \in \mathcal{S}$, $\lambda \mapsto f(u,\lambda)$ and $\lambda \mapsto r(u,\lambda)$ admit partial derivatives at $\lambda_0$ which satisfy:*

$$\frac{\partial f}{\partial \lambda}(u,\lambda_0) \quad = \quad -\left(E\left[X|T=u\right]f_0(u)\right)' \tag{8.1}$$

$$\frac{\partial r}{\partial \lambda}(u,\lambda_0) \quad = \quad -\left(E\left[DX|T=u\right]f_0(u)\right)' \tag{8.2}$$

**Proof:** let $X_{-m} = (X_1, ..., X_{m-1}, X_{m+1}..., X_p)$ and $f_{X_m|X_{-m}}(.,x)$ (resp. $F_{X_m|X_{-m}}(.,x)$) denote the density (resp. cumulative distribution function) of $X_m$ conditional on $X_{-m} = x$. Let also $\delta_k$ denote the vector of dimension $p$, with 1 at the k-th component and 0 elsewhere. We have

$$f(u,\lambda + t\delta_k) = \left|\begin{array}{ll} E\left[f_{X_m|X_{-m}}\left(\frac{u - X_{-m}'\lambda_{-m} - tX_k}{\lambda_m}, X_{-m}\right)\right] & \text{if } k \neq m, \\ E\left[f_{X_m|X_{-m}}\left(\frac{u - X_{-m}'\lambda_{-m}}{\lambda_m + t}, X_{-m}\right)\right] & \text{if } k = m. \end{array}\right.$$

Thus, by Assumption 3.2 and dominated convergence, $\lambda \mapsto f(u,\lambda)$ admits continuous partial derivatives. Now, let $F(.,\lambda)$ denote the cumulative distribution function of $X'\lambda$. We have,

$$F(u,\lambda + t\delta_k) = \left|\begin{array}{ll} E\left[F_{X_m|X_{-m}}\left(\frac{u - X_{-m}'\lambda_{-m} - tX_k}{\lambda_m}, X_{-m}\right)\right] & \text{if } k \neq m, \\ E\left[F_{X_m|X_{-m}}\left(\frac{u - X_{-m}'\lambda_{-m}}{\lambda_m + t}, X_{-m}\right)\right] & \text{if } k = m. \end{array}\right.$$

Thus, by Assumption 3.2 and dominated convergence, $\lambda \mapsto F(u, \lambda)$ admits continuous partial derivatives, and after some rearrangements,

$$\frac{\partial F}{\partial \lambda_k}(u, \lambda_0) = -E\left[X_k | T = u\right] f_0(u).$$

By Assumption 3.2 once more, $u \mapsto \partial F / \partial \lambda_k(u, \lambda_0)$ is continuously differentiable and

$$\frac{\partial^2 F}{\partial u \partial \lambda}(u, \lambda_0) = -\left(E\left[X | T = u\right] f_0(u)\right)'.$$

Then (8.1) follows from $\partial f / \partial \lambda = \partial^2 F / \partial \lambda \partial u = \partial^2 F / \partial u \partial \lambda$.

The proof of (8.2) is similar, except that we use $G_0(u, \lambda) = E(D\mathbb{1}\{X'\lambda \leq u\})$ instead of $F(u, \lambda)$. The partial derivatives of $\lambda \mapsto G_0(u, \lambda)$ exist and satisfy

$$\begin{aligned} \frac{\partial G_0}{\partial \lambda}(u, \lambda) &= -E\left(DX | T = u\right) f_0(u) \\ &= -S_{\eta_\Delta}(u + \delta_0) E\left(X | T = u\right) f_0(u). \end{aligned}$$

Then differentiability of $u \mapsto \partial G_0 / \partial \lambda(u, \lambda)$ stems from Assumptions 2.3 and 3.2. Equation (8.2) follows from the same argument as previously.

**Lemma 8.2** *Suppose that $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and Assumptions 3.2 and 3.6 hold. Then, for all $\mathcal{S}' \subsetneq \mathcal{S}$ and for all $\lambda_{u,n}$ such that $\sup_{u \in \mathcal{S}'} \|\lambda_{u,n} - \lambda_0\| = O_P(1/\sqrt{n})$, we have,*

$$\sup_{u \in \mathcal{S}'} |\widehat{q}(u, \lambda_{u,n}) - q_0(u)| = o_P(1) \tag{8.3}$$

$$\sup_{u \in \mathcal{S}'} \left\| \frac{\partial \widehat{q}}{\partial \lambda}(u, \lambda_{u,n}) - \frac{\partial q}{\partial \lambda}(u, \lambda_0) \right\| = o_P(1) \tag{8.4}$$

**Proof:** we first write

$$\sup_{u \in \mathcal{S}'} |\widehat{q}(u, \lambda_{u,n}) - q_0(u)| \leq \sup_{u \in \mathcal{S}'} |\widehat{q}(u, \lambda_{u,n}) - \widehat{q}(u, \lambda_0)| + \sup_{u \in \mathcal{S}'} |\widehat{q}(u, \lambda_0) - q_0(u)| \tag{8.5}$$

Let us first consider the the first term of the r.h.s. Since $|\widehat{q}(u, \lambda_{u,n})| \leq 1$, we have

$$\begin{aligned} \sup_{u \in \mathcal{S}'} |\widehat{q}(u, \lambda_{u,n}) - \widehat{q}(u, \lambda_0)| &= \sup_{u \in \mathcal{S}'} \frac{\left|(\widehat{r}(u, \lambda_{u,n}) - \widehat{r}(u, \lambda_0)) + \widehat{q}(u, \lambda_{u,n})(\widehat{f}(u, \lambda_0) - \widehat{f}(u, \lambda_{u,n}))\right|}{\widehat{f}(u, \lambda_0)} \\ &\leq \sup_{u \in \mathcal{S}'} \frac{1}{\widehat{f}(u, \lambda_0)} \left[|\widehat{r}(u, \lambda_{u,n}) - \widehat{r}(u, \lambda_0)| + \left|\widehat{f}(u, \lambda_{u,n}) - \widehat{f}(u, \lambda_0)\right|\right] \\ &\leq \frac{1}{\inf_{u \in \mathcal{S}'} \widehat{f}(u, \lambda_0)} \left[\sup_{u \in \mathcal{S}'} |\widehat{r}(u, \lambda_{u,n}) - \widehat{r}(u, \lambda_0)| \right. \\ &\qquad \left. + \sup_{u \in \mathcal{S}'} \left|\widehat{f}(u, \lambda_{u,n}) - \widehat{f}(u, \lambda_0)\right|\right]. \end{aligned} \tag{8.6}$$

Let us prove that

$$\sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\lambda_{u,n})-\widehat{f}(u,\lambda_0)\right| = o_P(1) \tag{8.7}$$

The proof for $\widehat{r}$ is similar. By Assumption 3.6, there exists $C_4>0$ such that $|K(u)-K(v)|\leq C_4|u-v|$. Thus,

$$\begin{aligned}
\left|\widehat{f}(u,\lambda_{u,n})-\widehat{f}(u,\lambda_0)\right| &\leq \frac{1}{nh_n}\sum_{i=1}^{n}\left|K\left(\frac{u-X_i'\lambda_{u,n}}{h_n}\right)-K\left(\frac{u-X_i'\lambda_0}{h_n}\right)\right| \\
&\leq \frac{C_4C_0\left\|\lambda_{u,n}-\lambda_0\right\|}{h_n^2} \\
&\leq \frac{C_4C_0\sup_{u\in\mathcal{S}'}\left\|\lambda_{u,n}-\lambda_0\right\|}{h_n^2} = O_p\left(\frac{1}{\sqrt{n}h_n^2}\right).
\end{aligned}$$

This establishes (8.7) since $nh_n^4\to\infty$. Because

$$\inf_{u\in\mathcal{S}'}\widehat{f}(u,\lambda_0)\geq -\sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\lambda_{u,n})-\widehat{f}(u,\lambda_0)\right| + \inf_{u\in\mathcal{S}'}f_0(u),$$

and because $\inf_{u\in\mathcal{S}'}f_0(u)>0$ by Assumption 3.2, we also have

$$\frac{1}{\inf_{u\in\mathcal{S}'}\widehat{f}(u,\lambda_0)} = O_p(1).$$

By (8.6), the first term of (8.5) tends to zero.

As for the second term, we can obtain the same decomposition as (8.6). Then Assumptions 3.2 and 3.6, and conditions on $h_n$ ensure that we can apply Lemma 8.10 of Newey & McFadden (1994), yielding $\sup_{u\in\mathcal{S}'}|\widehat{f}(u,\lambda_0)-f_0(u)| = o_P(1)$ and similarly for $\widehat{r}(.,\lambda_0)$. This establishes (8.3).

Now, let us turn to (8.4). We use the same decomposition as (8.5). First, let us establish that

$$\sup_{u\in\mathcal{S}'}\left|\frac{\partial\widehat{q}}{\partial\lambda}(u,\lambda_0)-\frac{\partial q}{\partial\lambda}(u,\lambda_0)\right| = o_P(1) \tag{8.8}$$

We have

$$\frac{\partial\widehat{q}}{\partial\lambda}(u,\lambda_0) = \frac{1}{\widehat{f}(u,\lambda_0)}\left[\frac{\partial\widehat{r}}{\partial\lambda}(u,\lambda_0)-\widehat{q}(u,\lambda_0)\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_0)\right].$$

and similarly for $\partial q/\partial\lambda(u,\lambda_0)$. Thus,

$$\begin{aligned}
&\frac{\partial\widehat{q}}{\partial\lambda}(u,\lambda_0)-\frac{\partial q}{\partial\lambda}(u,\lambda_0) \\
=\ &\frac{1}{\widehat{f}(u,\lambda_0)}\left\{\left[\frac{\partial\widehat{r}}{\partial\lambda}(u,\lambda_0)-\frac{\partial r}{\partial\lambda}(u,\lambda_0)\right]-\frac{\partial r}{\partial\lambda}(u,\lambda_0)\left[\frac{\widehat{f}(u,\lambda_0)-f_0(u)}{f_0(u)}\right]\right\} \\
&-\frac{\widehat{q}(u,\lambda_0)}{\widehat{f}(u,\lambda_0)}\left[\left(\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_0)-\frac{\partial f}{\partial\lambda}(u,\lambda_0)\right)-\frac{\partial f/\partial\lambda(u,\lambda_0)}{f_0(u)}\left(\widehat{f}(u,\lambda_0)-f_0(u)\right)\right] \\
&-\frac{\partial f/\partial\lambda(u,\lambda_0)}{f_0(u)}\left(\widehat{q}(u,\lambda_0)-q_0(u)\right).
\end{aligned}$$

By what precedes, $\inf_{u\in\mathcal{S}'}\widehat{f}(u,\lambda_0)$ tends in probability to $\inf_{u\in\mathcal{S}'} f_0(u)>0$, while $\sup_{u\in\mathcal{S}'}|\widehat{f}(u,\lambda_0)-f_0(u)|=o_P(1)$. Besides, $\widehat{q}(.,\lambda_0)$ is bounded by 1 and by Lemma 8.1, $\partial f/\partial\lambda(.,\lambda_0)$ is continuous on the compact set $\mathcal{S}$ and thus is bounded on this set. Thus, it suffices to prove that

$$\sup_{u\in\mathcal{S}'}\left|\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_0)-\frac{\partial f}{\partial\lambda}(u,\lambda_0)\right|=o_P(1) \tag{8.9}$$

and similarly for $r_0$. By Lemma 8.1, $u\mapsto\partial f/\partial\lambda(u,\lambda_0)$ is the derivative of $-E(X|T=u)f_0(u)$. As a consequence, we can apply Newey & McFadden (1994)'s Lemma 8.10, using as before Assumptions 3.2, 3.6, and conditions on $h_n$. This yields (8.9). The same reasoning applies to $r_0$, yielding (8.8).

Now, let us establish that

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{q}}{\partial\lambda}(u,\lambda_{u,n})-\frac{\partial\widehat{q}}{\partial\lambda}(u,\lambda_0)\right\|=o_P(1)$$

Using a similar decomposition as previously and the preceding results, it suffices to prove that

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_{u,n})-\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_0)\right\|=o_P(1) \tag{8.10}$$

and similarly for $\widehat{r}$. By Assumption 3.6, there exists $C_5>0$ such that $|K'(u)-K'(v)|\leq C_5|u-v|$. Thus,

$$\begin{aligned}\left\|\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_{u,n})-\frac{\partial\widehat{f}}{\partial\lambda}(u,\lambda_0)\right\| &\leq \frac{1}{nh_n^2}\sum_{i=1}^n\|X_i\|\left|K'\left(\frac{u-X_i'\lambda_{u,n}}{h_n}\right)-K'\left(\frac{u-X_i'\lambda_0}{h_n}\right)\right|\\ &\leq \frac{C_5C_0^2\,\|\lambda_{u,n}-\lambda_0\|}{h_n^3}=O_p\left(\frac{1}{\sqrt{n}h_n^3}\right).\end{aligned}$$

This proves (8.10) since $nh_n^6\to\infty$. The same reasoning applies to $\widehat{r}$. The result follows.

# 9 Appendix C: supplementary tables and figures

| Variables | $\zeta$ | $\beta_0$ | $\beta_1$ |
|---|---|---|---|
| Local average income | | | |
| Higher education graduates | 1.483*** (0.1) | 0 | 0.016*** (0.003) |
| High school graduates | -1 (0) | 0.019*** (0.004) | 0 |
| Secondary schooling track | | | |
| L | 8.757*** (0.476) | -0.066** (0.033) | -0.024 (0.019) |
| ES | 8.95*** (0.468) | -0.036 (0.034) | -0.019 (0.019) |
| S | 8.895*** (0.452) | -0.047 (0.033) | -0.029 (0.019) |
| Vocational | -28.502*** (0.493) | 0.234** (0.097) | -0.055 (0.071) |
| Technical | *Ref.* | *Ref.* | *Ref.* |
| Born abroad | 2.105*** (0.487) | -0.01 (0.016) | -0.004 (0.008) |
| Father born abroad | 1.151** (0.486) | -0.012 (0.01) | 0.008 (0.006) |
| Mother born abroad | 2.083*** (0.504) | -0.021* (0.012) | 0.008 (0.006) |
| Entering the labor market in 1998 (relative to 1992) | 6.48*** (0.416) | 0.113*** (0.024) | 0.144*** (0.015) |
| Male | 1.613*** (0.42) | 0.031*** (0.009) | 0.01** (0.004) |
| Father's profession | | | |
| Farmer | 1.238*** (0.467) | -0.004 (0.011) | 0.007 (0.007) |
| Tradesman | 1.477*** (0.455) | -0.009 (0.009) | -0.005 (0.005) |
| Executive | 5.236*** (0.452) | -0.034* (0.018) | 0.009 (0.011) |
| Intermediate occupation | 2.35*** (0.462) | -0.004 (0.011) | 0.004 (0.006) |
| Blue collar | -1.152*** (0.434) | 0.012* (0.007) | -0.009** (0.005) |
| Other | 0.965** (0.462) | -0.012 (0.009) | -0.012** (0.006) |
| White collar | *Ref.* | *Ref.* | *Ref.* |
| Mother's profession | | | |
| Farmer | -6.134*** (0.507) | 0.038* (0.023) | -0.027* (0.014) |
| Tradesman | -0.919* (0.5) | 0.012 (0.01) | -0.005 (0.006) |
| Executive | 0.973** (0.454) | -0.007 (0.011) | -0.008* (0.005) |
| Intermediate occupation | 0.624 (0.465) | 0.002 (0.01) | -0.003 (0.005) |
| Blue collar | -0.14 (0.461) | -0.001 (0.006) | 0.008* (0.004) |
| Other | 0.298 (0.464) | -0.007 (0.007) | -0.015*** (0.004) |
| White collar | *Ref.* | *Ref.* | *Ref.* |
| Age in 6th grade | | | |
| $\leq 10$ | 4.091*** (0.452) | -0.03* (0.017) | 0.005 (0.009) |
| 11 | *Ref.* | *Ref.* | *Ref.* |
| $\geq 12$ | -5.336*** (0.465) | 0.034* (0.019) | -0.013 (0.012) |
| Paris region | 1.886*** (0.467) | -0.001 (0.013) | 0.001 (0.005) |
| Vocational $\times$ ... | | | |
| Entering the labor market in 1998 | -0.018 (0.467) | -0.035*** (0.013) | -0.006 (0.012) |
| Male | 1.782*** (0.482) | -0.02* (0.01) | 0.015* (0.009) |
| Paris region | -4.175*** (0.498) | 0.022 (0.019) | -0.008 (0.014) |

Standard errors, presented in parentheses, were computed by bootstrap with 1,000 bootstrap sample replicates. Significativity levels: *** (1%), ** (5%) and * (10%).

Table 6: First step estimates

| Variable | Panel 1 | Panel 2 | Panel 3 |
|---|---|---|---|
| Constant ($\delta_0$) | -0.031 (0.134) | 0.089 (0.151) | 0.004 (0.119) |
| Local average income | | | |
| Higher education graduates | -0.012 (0.008) | -0.011 (0.008) | -0.012* (0.007) |
| Higher education graduation rate | -0.113* (0.064) | | |
| Secondary schooling track | | | |
| L | -0.122** (0.044) | -0.1** (0.041) | -0.123** (0.043) |
| ES | -0.155** (0.05) | -0.141** (0.049) | -0.151** (0.045) |
| S | -0.143** (0.044) | -0.124** (0.052) | -0.149** (0.043) |
| Vocational | 0.223 (0.144) | 0.251 (0.163) | 0.246* (0.139) |
| Technical | *Ref.* | *Ref.* | *Ref.* |
| Born abroad | -0.021 (0.016) | -0.023 (0.017) | -0.033* (0.018) |
| Father born abroad | -0.007 (0.015) | -0.013 (0.016) | -0.001 (0.01) |
| Mother born abroad | -0.01 (0.014) | -0.011 (0.014) | -0.01 (0.015) |
| Entering the labor market in 1998 (relative to 1992) | -0.08** (0.027) | -0.087** (0.031) | -0.09** (0.035) |
| Male | -0.057** (0.013) | -0.066** (0.016) | -0.051** (0.01) |
| Father's profession | | | |
| Farmer | -0.014 (0.013) | -0.014 (0.015) | -0.013 (0.014) |
| Tradesman | -0.03** (0.014) | -0.016* (0.01) | -0.024** (0.011) |
| Executive | -0.051** (0.022) | -0.049* (0.025) | -0.056** (0.023) |
| Intermediate occupation | -0.036** (0.014) | -0.04** (0.019) | -0.036** (0.013) |
| Blue collar | -0.002 (0.007) | 0 (0.008) | 0.001 (0.008) |
| Other | -0.017 (0.011) | -0.018 (0.012) | -0.018 (0.011) |
| White collar | *Ref.* | *Ref.* | *Ref.* |
| Mother's profession | | | |
| Farmer | 0.039 (0.025) | 0.038 (0.027) | 0.05 (0.031) |
| Tradesman | -0.004 (0.011) | 0 (0.012) | 0.001 (0.011) |
| Executive | -0.026* (0.015) | -0.025 (0.018) | -0.019* (0.01) |
| Intermediate occupation | -0.012 (0.009) | -0.014 (0.011) | -0.017 (0.012) |
| Blue collar | 0.014* (0.008) | 0.011 (0.009) | 0.012 (0.008) |
| Other | -0.016** (0.007) | -0.017** (0.008) | -0.014** (0.007) |
| White collar | *Ref.* | *Ref.* | *Ref.* |
| Age in 6th grade | | | |
| $\leq 10$ | -0.04* (0.023) | -0.043 (0.032) | -0.042** (0.021) |
| 11 | *Ref.* | *Ref.* | *Ref.* |
| $\geq 12$ | 0.055* (0.028) | 0.055* (0.03) | 0.053** (0.026) |
| Paris region | -0.022* (0.012) | -0.015 (0.013) | -0.033** (0.014) |
| Vocational × ... | | | |
| Entering the labor market in 1998 | 0.013 (0.021) | -0.002 (0.022) | 0.03 (0.019) |
| Male | 0.004 (0.014) | 0.019 (0.014) | 0.002 (0.013) |
| Paris region | 0.039* (0.022) | 0.025 (0.027) | 0.049* (0.026) |

In Panel 1, the higher education graduation rates are included in the estimation. In Panel 2 and 3, the streams of income were computed using ($\tau = 0.97, B = 25$) and ($\tau = 0.95, B = 30$) respectively. Standard errors, presented in parentheses, were computed by bootstrap with $1,000$ sample replicates. Significativity levels: $^{***}$ (1%), $^{**}$ (5%) and $^{*}$ (10%).

Table 7: Estimates of $\gamma$: robustness checks

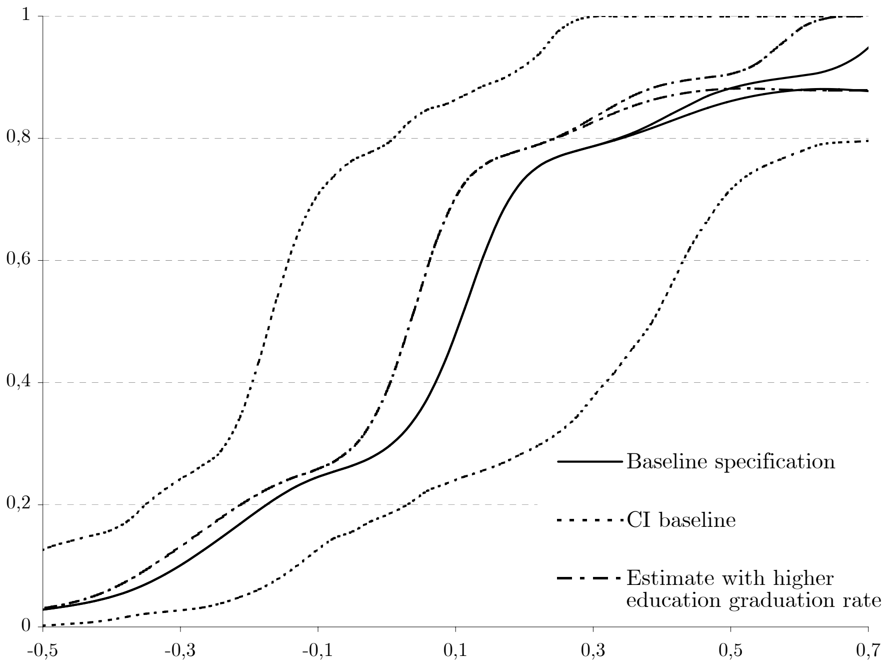


Figure 2: Returns of higher education including or not higher education graduation rate

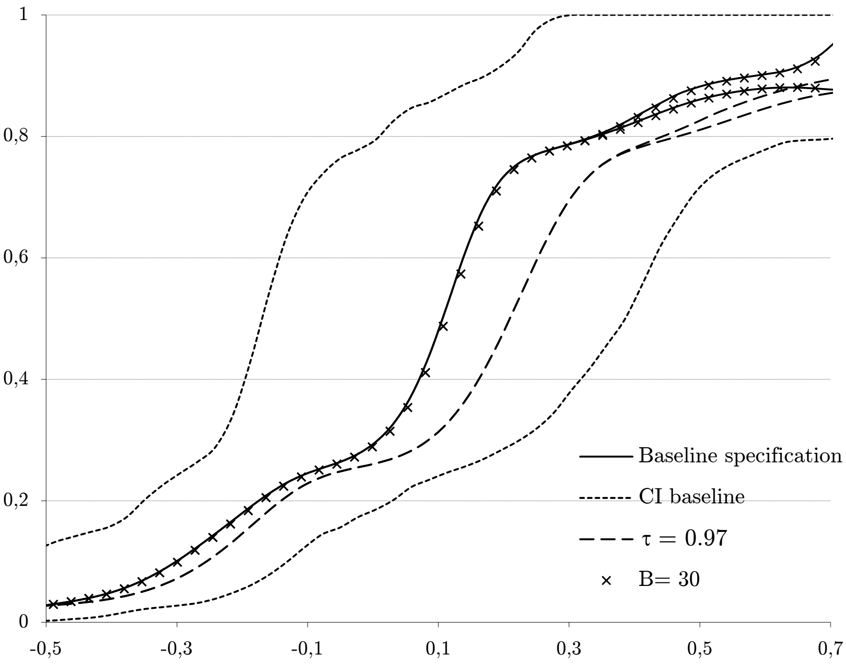


Figure 3: Returns of higher education under alternative computations of the stream of income

## References


Abbring, J. & van den Berg, G. (2003), 'The identifiability of the mixed proportional hzards competing risks model', *J.R. Statist. Soc. B* **65**, 701–710.

Ashenfelter, O. & Card, D. (1985), 'Using the longitudinal structure of earnings to estimate the effect of training programs', *Review of Economics and Statistics* **67**, 648–660.

Bayer, P. J., Khan, S. & Timmins, C. (2008), Nonparametric identification and estimation in a generalized roy model. NBER Working Paper.

Beffy, M., Fougère, D. & Maurel, A. (2009), 'Choosing the field of study in post-secondary education: Do expected earnings matter?', *IZA Discussion Paper No. 4127* .

Borjas, G. (1987), 'Self-selection and the earnings of immigrants', *The American Economic Review* **77**, 531–553.

Carneiro, P., Hansen, K. & Heckman, J. (2003), 'Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice', *International Economic Review* **44**, 361–422.

Carneiro, P. & Lee, S. (2009), 'Estimating distributions of potential outcomes using local instrumental variables with an application to change in college enrollment and wage inequality', *Journal of Econometrics* **149**, 191–208.

Cipollone, P. & Rosolia, A. (2007), 'Social interactions in high school: Lessons from an earthquake', *The American Economic Review* **97**, 948–965.

Coppejans, M. (2001), 'Estimation of the binary response model using a mixture of distributions estimator (mod)', *Journal of Econometrics* **102**, 231–269.

Cunha, F. & Heckman, J. (2007), 'Identifying and estimating the distributions of ex post and ex ante returns to schooling', *Labour Economics* **14**, 870–893.

d'Haultfoeuille, X. (2010), 'A new instrumental method for dealing with endogenous selection', *Journal of Econometrics* **154**, 1–15.

d'Haultfoeuille, X. & Maurel, A. (2009), Another look at the identification at infinity of sample selection models. IZA Discussion Paper No. 4334.

Dolton, P., Makepeace, G. & van der Klaauw, W. (1989), 'Occupational choice and earnings determination: The role of sample selection and non-pecuniary factors', *Oxford Economic Papers* **41**, 573–594.

Dustmann, C. & van Soest, A. (1998), 'Public and private sector wages of male workers in germany', *European Economic Review* **42**, 1417–1441.

Florens, J., Heckman, J. J., Meghir, C. & Vytlacil, E. (2008), 'Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects', *Econometrica* **76**, 1191–1206.

Ham, J. & LaLonde, R. (1996), 'The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training', *Econometrica* **64**, 175–205.

Heckman, J. J. (1974), 'Shadow prices, market wages, and labor supply', *Econometrica* **42**, 679–693.

Heckman, J. J. (1990), 'Varieties of selection bias', *The American Economic Review* **80**, 313–318.

Heckman, J. J. & Honore, B. (1989), 'The identifiability of competing risks models', *Biometrika* **76**, 325–330.

Heckman, J. J. & Honore, B. (1990), 'The empirical content of the Roy model', *Econometrica* **58**, 1121–1149.

Heckman, J. J. & Navarro, S. (2007), 'Dynamic discrete choice and dynamic treatment effects', *Journal of Econometrics* **136**, 341–396.

Heckman, J. & Vytlacil, E. (2005), 'Structural equations, treatment effects, and econometric policy evaluation', *Econometrica* **73**, 669–738.

Horowitz, J. L. (1998), *Semiparametric Methods in Econometrics*, Springer-Verlag.

Ichimura, H. (1993), 'Semiparametric least squares (SLS) and weighted SLS estimation of single-index models', *Journal of Econometrics* **58**, 71–120.

Klein, R. W. & Spady, R. H. (1993), 'An efficient semiparametric estimator for binary response models', *Econometrica* **61**, 387–421.

Kuruscu, B. (2006), 'Training and lifetime income', *The American Economic Review* **96**, 832–846.

Lalive, R. & Cattaneo, M. (2009), 'Social interactions and schooling decisions', *Review of Economics and Statistics* **91**, 457–477.

Lee, L. (1978), 'Unionism and relative wage rates: A simultaneous equations model with qualitative and limited dependent variables', *International Economic Review* **19**, 415–433.

Lee, S. (2006), 'Identification of a competing risks model with unknown transformations of latent failure times', *Biometrika* **93**, 996–1002.

Lee, S. & Lewbel, A. (2009), Identification in accelerated failure time competing risks models. Working Paper.

Linton, O. & Nielsen, J. P. (1995), 'A kernel method of estimating structured nonparametric regression based on marginal integration', *Biometrika* **82**, 93–100.

Manski, C. & Tamer, E. (2002), 'Inference on regressions with interval data on a regressor or outcome', *Econometrica* **70**, 519–546.

Newey, W. K. (2009), 'Two step series estimation of sample selection models', *The Econometrics Journal* **12**, S217–S229.

Newey, W. K. & McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. Engle & D. McFadden, eds, 'Hanbook of Econometrics', Vol. 4, Elsevier.

Peterson, A. (1976), 'Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks', *Proceedings of the National Academy of Science* **73**, 11–13.

Robinson, C. & Tomes, N. (1984), 'Union wage differentials in the public and private sectors: A simultaneous equations specification', *Journal of Labor Economics* **2**, 106–127.

Roy, A. D. (1951), 'Some thoughts on the distribution of earnings', *Oxford Economic Papers(New Series)* **3**, 135–146.

van den Berg, G. (1997), 'Association measures for durations in bivariate hazard rate models', *Journal of Econometrics* **79**, 221–245.

Willis, R. & Rosen, S. (1979), 'Education and self-selection', *Journal of Political Economy* **87**, S7–S36.

Yatchew, A. (1998), 'Nonparametric regression techniques in economics', *Journal of Economic Literature* **36**, 669–721.