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

Xavier D'Haultfœuille † Arnaud Maurel‡

JOB MARKET PAPER

September 2010

Abstract

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 earnings. This framework is well suited to various economic contexts, including educational and sectoral choices, or migration decisions. We show that the effects of covariates on earnings are identified under exclusion restrictions or at infinity. The non-pecuniary component can then be point or set identified without any other exclusion restriction, using the detailed structure of the model alone. Point identification is achieved if at least one covariate is continuous, while bounds are obtained otherwise. As a result, the distribution of the *ex ante* monetary returns can be point or set identified without any instrument. We propose a three-stage 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 and J24

Keywords: Roy model, nonparametric identification, schooling choices, *ex ante* returns to schooling.

*We thank Christian Belzil, Federico Bugni, Philippe Février, Marc Gurgand, Bo Honoré, Shakeeb Khan, Simon Lee, Thierry Magnac, Enno Mammen, Salvador Navarro, Jean-Marc Robin, Gerard van den Berg, participants at the 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), at the 2010 World Congress of the Econometric Society (Shanghai, August 2010) and in seminars at CREST (October 2009), Mannheim University (November 2009), Sciences Po (February 2010) and Duke University (September 2010) for helpful discussions and comments at various stages of this research.

†CREST (INSEE). E-mail address: xavier.dhaultfoeuille@ensae.fr.

‡Duke University, CREST and IZA. E-mail address: arnaud@amaurel.net

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 suggests that these non-pecuniary factors are indeed a key determinant of schooling decisions (see, e.g., Carneiro et al., 2003, or Beffy et al., 2010). 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., 2010).[1] Second, as emphasized by a recent stream of the literature on schooling choices (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.

In this paper, we explore what can be nonparametrically identified in a generalized Roy model including these two aspects, when relying on its detailed structure. We first develop two strategies for identifying the covariates effects on sector-specific earnings. The first one is based on exclusion restrictions. We require either a standard instrument, i.e. a variable affecting the selection probability but not the potential earnings, or sector-specific variables a la Heckman & Sedlacek (1985, 1990). The second strategy is to use an argument at infinity, relying on a recent result from a companion paper (D'Haultfoeuille & Maurel, 2009). We then turn to the identification of the non-pecuniary component. We

[1] We define *ex ante* monetary returns as the returns expected by the agent at the time of the choice.

show that without any other exclusion restriction, this component is point identified as soon as at least one covariate is continuous. When all covariates are discrete, we provide bounds on this non-pecuniary component. Noteworthy, our results are neither based 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 treatment effects. Even if no instrument is available, we obtain bounds on the distribution of the monetary benefits anticipated by the agents, which correspond in this setting to marginal treatment effects (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. (2010) 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. (2010) do not account for *ex ante* uncertainty, which may often be large. Second, their identification results are obtained under alternative assumptions.[2] 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, this condition may be restrictive, in particular when using log wages in utility functions, as for instance in Willis & Rosen (1979).[3] Bayer et al. (2010) alternatively prove identification under the assumption of independence between alternative-specific wages and the exclusion restriction that a variable affects the non-pecuniary factors of each choice alternative but not the wage distributions. When choosing not to use our argument at infinity, we also obtain identification of the model with similar exclusion restrictions, but the independence condition is not needed in our framework. This is convenient, since this assumption is restrictive, and much of the literature considering identification of Roy and the closely related competing risks models has produced alternative identification results without it (see, e.g., Heckman & Honoré, 1989, Heckman & Honoré, 1990, or Abbring & van den Berg, 2003).

Apart from identification, we also propose a three-stage semiparametric estimation procedure under an index restriction on the effects of the covariates. The first two stages allow to estimate the covariates effects on potential earnings and correspond to Newey's method

[2]Another difference is that their framework readily extends to three sectors or more, while ours does not.

[3]We also obtain identification of the full model without any exclusion restriction, using the aforementioned argument at infinity.

(2009) 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 simply amounts to estimate an instrumental linear model. The difference with a standard IV linear model is that both the dependent variable and one of the regressors have to be estimated, this involving in particular a non-parametric regression on generated covariates. We show that the corresponding estimator is root-n consistent and asymptotically normal. Monte Carlo simulations indicate that despite its multiple steps, the estimator performs fairly well in finite samples. They also show that when one of the regressors is restricted not to act on non-pecuniary factors, a constrained version of our estimator is substantially more accurate.

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. In a similar spirit as Carneiro et al. (2003), we estimate a model a la Willis & Rosen (1979), which is extended to account for non-pecuniary factors driving the attendance decision. We use respectively the local average incomes for high school and higher education graduates as sector-specific regressors, this yielding identification of the covariates effects on earnings. As may be expected, we cannot reject at the 10% level the assumption that the local average income for high school graduates only affects the probability of attendance through the *ex ante* returns to higher education. This allows us to apply our constrained estimator, leading to substantial gains of precision. Consistent with the recent evidence on this question, our results suggest that non-pecuniary factors are a key determinant of the decision to attend higher education. Comparing the influence of non-pecuniary factors with the one of *ex ante* monetary returns to education, we obtain that the median in the population of the non-pecuniary factors is about 2.5 times larger than the median of the *ex ante* returns to higher education, thus highlighting the major role played by non-pecuniary determinants in the decision to enroll in higher education. Noteworthy, unlike Carneiro et al. (2003), these results are not driven by a factor structure on the outcomes.

The remainder of the paper is organized as follows. Section 2 presents the extended Roy model which is considered throughout the paper and displays our identification results for the covariates effects on earnings and for the non-pecuniary component. Section 3 develops a semiparametric estimation procedure for this model, and proves the root-n consistency and asymptotic normality of the proposed estimators. Section 4 studies the finite-sample performances of the estimators by simulations. 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 as 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 earnings in sector k. These earnings 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 (η_0, η_1) are sector-specific productivity terms known by the agent at the time of the choice but unobserved by the econometrician. We 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, $X \perp\!\!\!\perp (\eta_0, \eta_1)$.*

The assumption that η_k is mean independent of X, i.e. $E(\eta_k|X)$ is constant, is without loss of generality. 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, 2009) 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.[4] 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) who posit a factor structure on the unobservables. Such a restriction is useful to identify the joint distribution of $(\eta_0, \eta_1, \nu_0, \nu_1)$, and thus to test for comparative advantage or to assess the importance of *ex post* uncertainty (see Cunha & Heckman, 2007). We do not consider these issues here.

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

[4]Part of the residual ν_k may correspond to a measurement error rather than an unexpected shock. We stick with the latter interpretation throughout the paper for convenience of exposition only.

$\mathcal{U}_k = E(Y_k|X, \eta_0, \eta_1) + G_k(X)$. Hence, $\mathcal{U}_k$ is assumed to be given by the sum of sector-specific expected earnings $E(Y_k|X, \eta_0, \eta_1)$ and the non-pecuniary component associated with sector k, $G_k(X)$, which is supposed to depend on the covariates X. Along with the covariates X, the econometrician observes the chosen sector D, which therefore 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 earnings 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 effects. Second, our approach does not require any exclusion restriction between the selection equation and the potential earnings. As is shown in the following, the availability of sector-specific regressors allows to identify the covariates effects and the non-pecuniary component. Identification can also be achieved without any exclusion restriction, using arguments at infinity.

We maintain the following assumptions subsequently.

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

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

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

Assumption 2.2 is an innocuous normalization which stems from the fact that adding a constant to ψ_k and subtracting it to η_k does not modify the model. Assumption 2.3 is a technical condition which is usual in competing risks or Roy models (see, e.g., Heckman & Honoré, 1990, or Lee, 2006).

2.2 Identification of the covariates effects on earnings

Before detailing our key result on the identification of the non-pecuniary component G, we present in this subsection two alternative strategies to recover (ψ_0, ψ_1). The first is rather standard and relies on exclusion restrictions, in a similar spirit as in, e.g., Heckman & Honoré (1990). The second yields identification at infinity, and presents the advantage of not requiring any exclusion restriction. The first strategy is based on the following assumption.

Assumption 2.4 *(Exclusion restrictions) ψ_0 (resp. ψ_1) depends only on $\widetilde{X}_0 \subset X$ (resp. on $\widetilde{X}_1 \subset X$). Moreover, $\widetilde{X}_0$ (resp. $\widetilde{X}_1$) and $P(D = 1|X)$ are measurably separated, that is, any function of $\widetilde{X}_0$ (resp. of $\widetilde{X}_1$) almost surely equal to a function of $P(D = 1|X)$ is almost surely constant.*

The first part of Assumption 2.4 covers two rather different situations. The first one is when $X = (\widetilde{X}_0, Z)$ and $\widetilde{X}_1 = \widetilde{X}_0$. This corresponds to the standard instrumental setting in sample selection models, where the instrument Z affects the probability of selection but not the potential outcomes. In our framework, Z would be a determinant of the non-pecuniary component but not of the potential earnings. The second situation corresponds to the case where $X = (X_0, X_1, X_c)$, $\widetilde{X}_0 = (X_0, X_c)$ and $\widetilde{X}_1 = (X_1, X_c)$. This occurs in the presence of sector-specific regressors. In this case, no exclusion restriction between the non-pecuniary factors and the potential earnings is required. This kind of exclusion restrictions was previously used in particular by Heckman & Sedlacek (1985, 1990) when estimating parametrically a multiple-sector Roy model of self-selection in the labor market. We also rely on sector-specific regressors later on in our application.

Intuitively, 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 when, considering the two cases above, Z or (X_0, X_1) is continuous (see Florens et al., 2008, for sufficient conditions in this case). However, it may not hold when Z (or (X_0, X_1)) is discrete. As an illustration, consider a standard instrumental setting where $\widetilde{X}_0$ and Z are

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

binary and let $P_{ij} = P(D = 1|\widetilde{X}_0 = i, Z = j)$ for $i, j \in \{0, 1\}$. Then, provided that P_{10} and P_{11} do not belong to $\{P_{00}, P_{01}\}$, there exists a function h such that $h(P_{00}) = h(P_{01})$ and $h(P_{10}) = h(P_{11})$ but $h(P_{00}) \neq h(P_{10})$. In this case, the function g defined by $g(0) = h(P_{00})$ and $g(1) = h(P_{10})$ is not constant. As a result, $\widetilde{X}_0$ and $P(D = 1|X)$ are not measurably separated.

Given the preceding exclusion restrictions and the additive decomposition assumption, it is possible to identify ψ_0 and ψ_1 up to location parameters. Then full identification stems from the normalization of Assumption 2.2. Note that Theorem 2.1 below 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 ψ_0 and ψ_1 are identified.*

Alternatively, ψ_0 and ψ_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 ε_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 of the shocks (ν_0, ν_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 η_0 and η_1. When $(\eta_0, \eta_1, \nu_0, \nu_1)$ is gaussian for instance, one can show that it is equivalent to $\mathrm{cov}(\eta_0, \eta_1) < \min(V(\eta_0), V(\eta_1))$. 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 ψ_0 and ψ_1 are identified.*

[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.

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 & Honoré (1989) and Abbring & van den Berg (2003) in the related competing risks model. Nevertheless, their results cannot 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). Besides, this identification strategy does not rely on any support condition on X. In particular, it may be applied even if X is discrete, while the first proposed strategy generally fails in that case. 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 suppose for that purpose that one of the two frameworks displayed above can be used to identify (ψ_0, ψ_1), and that Assumption 2.3 holds. Letting $T(X) = \psi_0(X) - \psi_1(X)$, we start from the following observations:

$$E[D\eta_\Delta | X] = E\left[\mathbb{1}\{\eta_\Delta \geq T(X) + G(X)\}\eta_\Delta\right] = \int_{T(X)+G(X)}^{\infty} u f_{\eta_\Delta}(u) du, \tag{2.3}$$

$$E[D|X] = \int_{T(X)+G(X)}^{\infty} f_{\eta_\Delta}(u) du. \tag{2.4}$$

We first suppose that at least one of the components X_j of X, say X_1, is continuous, and impose the following regularity condition.

Assumption 2.6 *X_1 is continuous and $(T + G)(.)$ is differentiable on its support with respect to x_1.*

Let $X_{-1} = (X_2, ...)$ denote the other elements of X. Under Assumptions 2.3 and 2.6, the functions $q_0(x_1, x_{-1}) = E(D|X_1 = x_1, X_{-1} = x_{-1})$ and $E[D\eta_\Delta | X_1 = x_1, X_{-1} = x_{-1}]$ are

[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 to generalized Roy models either.

differentiable with respect to x_1. Besides, letting $x = (x_1, x_{-1})$, we obtain

$$\frac{\partial E[D\eta_\Delta | X_1 = x_1, X_{-1} = x_{-1}]}{\partial x_1} = (T(x) + G(x))\frac{\partial q_0}{\partial x_1}(x_1, x_{-1}).$$

Now, the definition of ν_i and the law of iterated expectations yield $E(\nu_k | D = k, X) = 0$. As a result, letting $\varepsilon = D\varepsilon_1 + (1 - D)\varepsilon_0$, we get

$$\begin{aligned} E(\varepsilon | X) &= E[D\varepsilon_1 + (1 - D)\varepsilon_0 | X] \\ &= E[D\eta_1 + (1 - D)\eta_0 | X] \\ &= E[D\eta_\Delta | X] + E[\eta_0]. \end{aligned} \tag{2.5}$$

Thus, letting $g_0(x_1, x_{-1}) = E(\varepsilon | X_1 = x_1, X_{-1} = x_{-1})$, we obtain

$$\frac{\partial g_0}{\partial x_1}(x_1, x_{-1}) = (T(x) + G(x))\, \frac{\partial q_0}{\partial x_1}(x_1, x_{-1}). \tag{2.6}$$

Because $\varepsilon = Y - \psi_D(X)$ is identified, g_0 and q_0 are identified and we can use Equation (2.6) to recover $G(.)$. The only exception is when $\frac{\partial q_0}{\partial x_1}$ is identically equal to zero, a case which is ruled out by Assumption 2.7 below. Theorem 2.3 shows that, under this condition, $G(.)$ is identified.[9]

Assumption 2.7 *For all $u \in \mathbb{R}$, $f_{\eta_\Delta}(u) > 0$ and for all x_{-1} in the support of X_{-1}, the set $\{x_1 / \frac{\partial(T+G)}{\partial x_1}(x_1, x_{-1}) \neq 0\}$ is not empty.*

Theorem 2.3 *Suppose that $T(.)$ is identified and Assumptions 2.3, 2.6 and 2.7 hold. Then $G(.)$ is identified.*

Now consider the case where X has a discrete distribution and takes M values $x_1, ..., x_M$. Then one cannot take the derivative of g_0 and q_0 anymore. However, the strategy above can be adapted to yield bounds on G. First, note that $P(D = 0 | X = x) = F_{\eta_\Delta}(T(x) + G(x))$, with F_{η_Δ} denoting the cumulative distribution function of η_Δ. This equality implies that we can sort the x_i's so that $T(x_1) + G(x_1) < ... < T(x_M) + G(x_M)$.[10] This provides a first

[9] If the second condition of Assumption 2.7 fails to hold, $\frac{\partial G}{\partial x_1}$ is still identified, as it is equal to $-\frac{\partial T}{\partial x_1}$ in this case. Note also that because Assumption 2.7 implies that $\frac{\partial q_0}{\partial x_1}$ is not identically equal to zero, this restriction can be tested in the data.

[10] This is without loss of generality. In case of ties between $T(x_i) + G(x_i)$ and $T(x_{i+1}) + G(x_{i+1})$, one may remove x_{i+1} from the set of x's. Then the bounds on $G(x_{i+1})$ follow directly from those on $G(x_i)$.

set of inequalities on $(G(x_1), ..., G(x_M))$. Besides, letting $i < j$, we have,

$$\begin{aligned} & \sum_{k=i}^{j-1} \left[T(x_{k+1}) + G(x_{k+1})\right]\left[q_0(x_{k+1}) - q_0(x_k)\right] \\ \leq \;& g_0(x_j) - g_0(x_i) = -\int_{T(x_i)+G(x_i)}^{T(x_j)+G(x_j)} u f_{\eta_\Delta}(u) du \\ \leq \;& \sum_{k=i}^{j-1} \left[T(x_k) + G(x_k)\right]\left[q_0(x_{k+1}) - q_0(x_k)\right]. \end{aligned}$$

These inequalities provide supplementary conditions for $(G(x_1), ..., G(x_M))$. Note that we only get an upper bound for $G(x_1)$ and a lower bound for $G(x_M)$, but both for $G(x_2), ..., G(x_{M-1})$.

2.4 Distribution of treatment effects

We now turn to the identification of the distribution of the *ex ante* treatment effect, $\Delta = E(Y_1 - Y_0|X, \eta_0, \eta_1)$. The *ex ante* treatment effect is meaningful since it corresponds to what agents act on (see Cunha & Heckman, 2007). Besides, it corresponds to the *ex post* treatment effect if (i) agents perfectly observe or anticipate 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. To identify the *ex ante* treatment effect, we start from

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

This shows that F_{η_Δ} is identified over the support of $T(X) + G(X)$. Now, the cumulative distribution function of Δ writes

$$\begin{aligned} F_\Delta(u) &= E\left[P(\eta_\Delta \leq u + T(X)|X)\right] \\ &= E\left[F_{\eta_\Delta}(u + T(X))\right]. \end{aligned}$$

Hence, we can identify $F_\Delta(u)$ for all u such that the support of $u + T(X)$ is included in the support of $T(X) + G(X)$. In particular, the complete distribution of the *ex ante* treatment effect Δ is identified as soon as $T(X) + G(X)$ 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, by integrating the *ex ante* treatment effect over the distribution of η_Δ. 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 $T(X)+G(X)$ (resp. of $P(D=0|X)$), we have, by the monotonicity of F_{η_Δ}, $F_\Delta(u) \in [\underline{F_\Delta}(u), \overline{F_\Delta}(u)]$, where

$$\begin{aligned}
\underline{F_\Delta}(u) &= E\left(F_{\eta_\Delta}(u+T(X))\,\mathbb{1}\{u+T(X)\in[\underline{M},\overline{M}]\}\right) \\
&\quad +\overline{P}\times P(u+T(X)>\overline{M}) + 0\times P(u+T(X)\leq \underline{M}), \qquad (2.7)\\
\overline{F_\Delta}(u) &= E\left(F_{\eta_\Delta}(u+T(X))\,\mathbb{1}\{u+T(X)\in[\underline{M},\overline{M}]\}\right) \\
&\quad +1\times P(u+T(X)>\overline{M}) + \underline{P}\times P(u+T(X)\leq \underline{M}). \qquad (2.8)
\end{aligned}$$

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

$$F_{\Delta|D=1}(u) = \frac{E\{(F_{\eta_\Delta}(u+T(X)) - P(D=0|X))\times \mathbb{1}\{G(X)\leq u\}\}}{P(D=1)}. \qquad (2.9)$$

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

$$\begin{aligned}
\Delta^{MTE}(x,p) &= E(Y_1-Y_0|X=x, S_{\eta_\Delta}(\eta_\Delta)=p) \\
&= \psi_1(x)-\psi_0(x)+S_{\eta_\Delta}^{-1}(p)
\end{aligned}$$

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 $\widetilde{x}$ in the support of X such that $S_{\eta_\Delta}^{-1}(p) = (\psi_0-\psi_1+G)(\widetilde{x})$.

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 estimators of $\psi_0(.), \psi_1(.)$ and $G(.)$. More precisely, we consider generalized Roy models with a linear index structure of the form:[11]

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

In this setting, the non-pecuniary component $G(X)$ is of the form $\delta_0+X'\gamma_0$. Let γ_{0j} (resp. β_{0j}, β_{1j}) denote the j-th component of γ_0 (resp. β_0, β_1). We also impose the following conditions.

[11]We suppose without loss of generality that the constant is not included in X, so that ε_0 and ε_1 do not necessarily have mean zero.

Assumption 3.1 *(Exclusion restrictions) There exists* j_1 *and* j_2 *such that* $\beta_{0j_1} = \beta_{1j_2} = 0$, $\gamma_{0j_1} \neq \beta_{1j_1}$ *and* $\gamma_{0j_2} \neq -\beta_{0j_2}$.

Assumption 3.2 *(Regularity of* X*) The support of* X *is bounded. For all* x_{-1} *in the support of* X_{-1}*, the distribution of* X_1 *conditional on* $X_{-1} = x_{-1}$ *admits a continuously differentiable and positive density on its support, which is a compact interval independent of* x_{-1}*. Besides,* $\beta_{11} - \beta_{01} - \gamma_{01} \neq 0$*. Moreover, for all* j, $t \mapsto E(X_j|X'(\beta_1 - \beta_0 - \gamma_0) = t)$ *is continuously differentiable. Finally, the support of* $X'(\beta_1 - \beta_0 - \gamma_0)$ *is an interval.*

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).

Assumption 3.1 corresponds, in this semiparametric framework, to Assumption 2.4. The case where $j_1 = j_2$ corresponds to the standard instrumental variable setting of sample selection models, while $j_1 \neq j_2$ applies when some covariates are sector-specific. Assumption 3.2 corresponds to Assumptions 2.6 and 2.7. It ensures that at least one covariate is continuous and has a nonzero effect on D (because $\beta_{11} - \beta_{01} - \gamma_{01} \neq 0$). As shown in Theorem 2.3, this condition is sufficient to provide point identification of $G(.)$. We also impose the support of $X'(\beta_1 - \beta_0 - \gamma_0)$ to be an interval. This condition is sufficient to point identify the single index model on D (see, e.g., Horowitz, 1998) that corresponds to our first step estimator described below.

Let us assume, without loss of generality, that $\beta_{11} - \beta_{01} - \gamma_{01}$ is strictly positive. We define $\zeta_0 = (\beta_1 - \beta_0 - \gamma_0)/(\beta_{11} - \beta_{01} - \gamma_{01})$ (so that $\zeta_{01} = 1$) and $\widetilde{\eta}_\Delta = (\eta_\Delta - \delta_0)/(\beta_{11} - \beta_{01} - \gamma_{01})$. We propose a three-stage estimation procedure of the model. The first and second stages of our procedure rely on the fact that we can rewrite the model as

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

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 $X'\zeta_0$.[12] Besides, by Assumption 3.1, X_{j_1} (resp. X_{j_2}) affects selection since $\zeta_{0j_1} \neq 0$ (resp. $\zeta_{0j_2} \neq 0$) but not the potential earnings 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 ζ_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 β_0 and β_1 by series estimator, and

[12] 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, ε_k is not independent of X because ν_k is not.

we suppose that it satisfies Assumption 3.5. This condition can be obtained under more primitive assumptions (see Newey, 2009, p. S227).

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

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

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

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

Now let us turn to the estimation of (δ_0,γ_0). First note that actually, it suffices to estimate δ_0 and $\alpha_0\equiv\beta_{01}-\beta_{11}+\gamma_{01}$, since $\gamma_0=\beta_1-\beta_0+\alpha_0\zeta_0$. Equations (2.3), (2.4) and (2.5) applied to the current index model show that $E(D|X)$ and $E(\varepsilon|X)$ only depend on $U\equiv X'\zeta_0$. Letting, with a slight abuse of notation, $q_0(u)=E(D|U=u)$ and $g_0(u)=E(\varepsilon|U=u)$, we have, similarly to Equation (2.6),

$$g_0'(U)=q_0'(U)(\delta_0+\alpha_0 U). \tag{3.3}$$

Integrating (3.3) between u_0 in the support of U and U, we obtain:

$$g_0(U)=\widetilde{\lambda}_0+q_0(U)\delta_0+\left[\int_{u_0}^U uq_0'(u)du\right]\alpha_0,$$

where $\widetilde{\lambda}_0$ is the constant of integration. An integration by part yields

$$g_0(U)=\lambda_0+q_0(U)\delta_0+\left[q_0(U)U-\int_{u_0}^U q_0(u)du\right]\alpha_0, \tag{3.4}$$

where $\lambda_0=\widetilde{\lambda}_0-u_0q_0(u_0)\alpha_0$. In other terms,

$$\varepsilon=\lambda_0+D\delta_0+\left[DU-\int_{u_0}^U q_0(u)du\right]\alpha_0+\xi,\quad E(\xi|X)=E(\xi|U)=0. \tag{3.5}$$

Let $\theta_0=(\lambda_0,\delta_0,\alpha_0)'$, $V=DU-\int_{u_0}^U q_0(u)du$ and $W=(1,D,V)'$, so that $\varepsilon=W'\theta_0+\xi$. We estimate θ_0 with an IV estimator which, for technical reasons, includes some trimming. We consider (unfeasible) instruments of the kind $Z=\mathbb{1}\{X\in\mathcal{X}\}h(U)$, where $h(U)=$

$(1, h_1(U), h_2(U))' \in \mathbb{R}^3$ and $\mathcal{X}$ is a set included in the support of X and such that $\{x'\zeta_0, x \in \mathcal{X}\}$ is a closed interval strictly included in the interior of the support of U. Then $\theta_0 = E(ZW')^{-1}E(Z\varepsilon)$, and we estimate it by

$$\widehat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{Z}_i \widehat{W}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{Z}_i \widehat{\varepsilon}_i\right),$$

where $\widehat{\varepsilon}_i = Y_i - X_i'(D_i\widehat{\beta}_1 + (1 - D_i)\widehat{\beta}_0)$, $\widehat{W}_i = (1, D_i, \widehat{V}_i)'$ and

$$\begin{aligned} \widehat{V}_i &= D_i\widehat{U}_i - \int_{u_0}^{\widehat{U}_i} \widehat{q}(u, \widehat{\zeta})du, \\ \widehat{Z}_i &= \mathbb{1}\{X_i \in \mathcal{X}\}h\left(\widehat{U}_i\right). \end{aligned}$$

Finally, $\widehat{U}_i = X_i'\widehat{\zeta}$ and

$$\widehat{q}(u, \zeta) = \frac{\sum_{i=1}^{n} D_i K\left(\frac{u - X_i'\zeta}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{u - X_i'\zeta}{h_n}\right)}. \tag{3.6}$$

where $K(.)$ is a kernel function and h_n the bandwidth 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) $h_k(.)$ is twice differentiable and $|h_k''|$ is bounded for $k \in \{1, 2\}$.*

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 $\widehat{Z}_i - Z_i$ is small for large sample sizes, 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*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0,\ E(ZW')^{-1}V(Z\xi + \Omega_{11} + \Omega_{21})E(WZ')^{-1}\right),$$

where Ω_{11} is defined by Equation (7.8) in Appendix A and

$$\Omega_{21} = \alpha_0 Z(1 - F_0(U))\mathbb{1}\{U \geq u_0\}(D - q_0(U))/f_0(U),$$

$F_0(.)$ and $f_0(.)$ denoting respectively the cumulative distribution function and the density of U.

Theorem 3.1 establishes the asymptotic normality of $\theta_0 = (\lambda_0, \delta_0, \alpha_0)'$ and shows that its asymptotic variance depends on the three variables Ω_{11}, Ω_{21} and $Z\xi$. 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 $q_0(.)$ in $\widehat{V}_i$. The third one corresponds to the moment estimation of the linear instrumental model (3.5) in the last step. This theorem guarantees that $\widehat{\delta}$ is root-n consistent and asymptotically normal. As the proof of the theorem shows, $\widehat{\theta}$ can be linearized. Thus, by Assumptions 3.4 and 3.5, the estimator of γ_0, $\widehat{\gamma} = \widehat{\beta}_1 - \widehat{\beta}_0 + \widehat{\alpha}\widehat{\zeta}$, is also root-n consistent and asymptotically normal.

Although δ_0 and γ_0 are identified without any exclusion restriction, imposing some restriction on γ_0 may still be useful to improve the accuracy of the estimators. Suppose that, e.g., X_1 is excluded from the non-pecuniary component, so that $\gamma_{01} = 0$.[13] In this case, we get from the second stage $\alpha_0 = \beta_{01} - \beta_{11}$, and thus γ_0 is identified directly by $\beta_1 - \beta_0 + \alpha_0\zeta_0$. Hence, γ_0 can be estimated using only the first two steps, resulting in general in accuracy gains (see our Monte Carlo simulations and application below for evidence on this point). The third stage boils down to estimating δ_0 only, through the instrumental linear model

$$\varepsilon - \left[DU - \int_{u_0}^{U} q_0(u)du\right]\alpha_0 = \lambda_0 + D\delta_0 + \xi, \quad E(\xi|X) = E(\xi|U) = 0, \tag{3.7}$$

where α_0 in the left hand side can now be estimated by $\widehat{\beta}_{01} - \widehat{\beta}_{11}$. One can show that the corresponding estimator is also asymptotically normal.[14]

4 Monte Carlo simulations

In this section, we investigate the finite-sample performance of our semiparametric estimation procedure by simulating the following model with sector-specific variables:

$$\begin{aligned}
Y_{0i} &= X_{2i}\beta_{02} + X_{3i}\beta_{03} + \eta_{0i} + \nu_{0i} \\
Y_{1i} &= X_{1i}\beta_{11} + X_{3i}\beta_{13} + \eta_{1i} + \nu_{1i} \\
D_i &= \mathbb{1}\{-\delta_0 + X_{1i}(\beta_{11} - \gamma_{01}) + X_{2i}(-\beta_{02} - \gamma_{02}) + X_{3i}(\beta_{13} - \beta_{03} - \gamma_{03}) + \eta_{1i} - \eta_{0i} > 0\}.
\end{aligned}$$

The true values of the parameters are $\beta_{02} = \beta_{03} = 1$, $\beta_{11} = 2$, $\beta_{13} = 0.5$, $\gamma_{01} = -0.5$, $\gamma_{02} = 0.5$, $\gamma_{03} = -0.8$ and $\delta_0 = 0.8$, so that Assumption 3.1 is satisfied with $j_1 = 1$ and $j_2 = 2$. 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 $p = 0.5$.

[13] Of course, such an assumption can be tested by estimating the unrestricted model.

[14] The proof is very close to the one of Theorem 3.1 and is available from the authors upon request.

We let $(\eta_{0i}, \eta_{1i})'$ be joint normal, with zero mean and a variance Σ 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 zero mean and a conditional matrix 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-stage estimation procedure detailed in Section 3. We estimate in the first stage $\zeta_0 = (\beta_1 - \beta_0 - \gamma_0)/(\beta_{11} - \gamma_{01})$ by Klein & Spady's (1993) semiparametric efficient estimator, with an adaptive gaussian kernel and local smoothing. In the second stage, we implement Newey's (2009) method in order to estimate separately β_0 and β_1. The series estimator of the selection correction term was computed using the inverse Mills ratio transform (see Newey, 2009, 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 estimators for δ_0 and γ_0 with the quartic kernel suggested in Section 3 and a bandwidth $h_n = 0.5\sigma(\widehat{U})n^{-1/7}$, where $\sigma(\widehat{U})$ is the estimated standard deviation of $\widehat{U}$. We choose the functions $h_1(x) = \Phi(\widehat{a}_0 + \widehat{a}_1 x)$ and $h_2(x) = xh_1(x) - \int_{\widehat{u}_0}^{x} \widehat{q}(u, \widehat{\zeta})du$ for the instruments, where $\Phi(.)$ denotes the normal cumulative distribution, $(\widehat{a}_0, \widehat{a}_1)$ is the probit estimator of D on $(1, \widehat{U})$ and $\widehat{u}_0$ is the sample minimum of $\widehat{U}$.[15] Finally, no trimming was performed since it did not improve the accuracy of the estimators 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 Panel A of Table 1 below, which reports for each parameter its bias, standard deviation and root mean squared error (RMSE). The results indicate that our procedure performs reasonably well in this context. As expected given the sequential structure of the proposed estimation procedure, the non-pecuniary components δ_0 and γ_0 are less precisely estimated than β_0 and β_1. Their estimators also display a substantial bias until $n = 1,000$, but this bias seems to decrease quickly for larger samples.

[15] For the sake of simplicity, we suppose in Section 3 that the functions $h(.)$ are known to the econometrician. Assuming alternatively that these functions have to be estimated, as is the case here, does not affect the root-n consistency and asymptotic normality of the estimators.

		Panel A			Panel B			Panel C		
n	Coeff.	Bias	Std dev	RMSE	Bias	Std dev	RMSE	Bias	Std dev	RMSE
500	β_{02}	-0.023	0.132	0.134	-0.011	0.118	0.118	-0.011	0.118	0.119
	β_{03}	-0.004	0.263	0.263	-0.003	0.214	0.214	0.007	0.221	0.222
	β_{11}	0.030	0.162	0.165	0.018	0.147	0.148	0.012	0.150	0.150
	β_{13}	0.000	0.193	0.193	-0.002	0.199	0.199	-0.016	0.208	0.208
	γ_{01}	0.024	1.631	1.631	0.057	1.144	1.145		*(not estimated)*	
	γ_{02}	-0.128	0.918	0.926	-0.085	0.842	0.846	-0.057	0.186	0.194
	γ_{03}	0.037	0.425	0.426	0.020	0.386	0.387	-0.022	0.361	0.362
	δ_0	0.153	0.876	0.889	0.030	0.763	0.763	0.098	0.563	0.572
1,000	β_{02}	-0.021	0.093	0.095	-0.012	0.083	0.084	-0.007	0.086	0.087
	β_{03}	-0.002	0.176	0.176	0.006	0.156	0.156	0.009	0.150	0.150
	β_{11}	0.021	0.112	0.114	0.012	0.102	0.103	0.011	0.105	0.106
	β_{13}	0.007	0.133	0.133	-0.005	0.145	0.145	-0.001	0.143	0.143
	γ_{01}	-0.016	1.137	1.137	0.061	0.792	0.794		*(not estimated)*	
	γ_{02}	-0.086	0.643	0.649	-0.087	0.588	0.594	-0.042	0.133	0.139
	γ_{03}	0.020	0.289	0.290	0.007	0.275	0.275	0.001	0.239	0.239
	δ_0	0.148	0.591	0.609	0.038	0.520	0.521	0.070	0.412	0.418
2,000	β_{02}	-0.021	0.067	0.070	-0.006	0.061	0.061	-0.009	0.060	0.060
	β_{03}	0.004	0.118	0.118	-0.004	0.110	0.110	-0.003	0.114	0.114
	β_{11}	0.014	0.078	0.079	0.010	0.073	0.073	0.013	0.072	0.073
	β_{13}	-0.005	0.092	0.092	0.001	0.102	0.102	-0.005	0.101	0.101
	γ_{01}	0.014	0.799	0.799	0.028	0.574	0.574		*(not estimated)*	
	γ_{02}	-0.073	0.461	0.467	-0.053	0.425	0.428	-0.032	0.089	0.095
	γ_{03}	0.011	0.195	0.196	0.013	0.186	0.187	0.004	0.175	0.175
	δ_0	0.088	0.415	0.424	0.032	0.374	0.376	0.047	0.278	0.282

Note: Panel A corresponds to the unconstrained model, while in Panel B and Panel C, $\gamma_{01} = 0$. In Panel B we suppose that the econometrician ignores this restriction, so that (δ_0, γ_0) are estimated with (3.5). In panel C, the econometrician knows it, and estimates are based on (3.7). The results were obtained with 1,000 simulations for each sample size.

Table 1: Monte Carlo simulations

We also investigate the effect of using an exclusion restriction on the non-pecuniary component on the finite-sample performances of the estimators. For that purpose, we consider the same specification as previously with the exception that $\gamma_{01} = 0$, and compare estimates obtained when this restriction is known by the econometrician and when it is not.

As explained in the previous section, we can recover γ_0 in the former case with the first step estimates alone, and use Equation (3.7) to estimate δ_0 only. The properties of the unconstrained and constrained estimators are displayed respectively in Panel B and C of Table 1. Overall, using an exclusion restriction on $G(.)$ leads to a substantial improvement in the performances of the estimators of γ_0 and δ_0. In particular, the RMSE for γ_{02} decreases by about 80% between the two specifications. Although still estimated in the last step, the estimator for δ_0 performs significantly better than in the unconstrained setting, with the RMSE decreasing by about 25%.

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 factors and monetary returns to education in the decision to attend higher education in France. We first briefly present in Subsection 5.1 the underlying 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 discuss 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 non-pecuniary consumption value of schooling, in a semiparametric setting.[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 ($k = 1$).[17] They are supposed to make their decision $D \in \{0, 1\}$ by comparing the expected discounted

[16] On a related ground, Carneiro et al. (2003) also estimate a generalization of the Willis & Rosen's model accounting for non-pecuniary factors affecting the decision to attend college. Nevertheless, they rely on a completely different framework based on factor loadings, which is quite demanding in terms of identifying conditions. Apart from the existence of regressors entering the selection equation only, 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) and Carneiro et al. (2010) who estimate on the same dataset a semiparametric reduced-form model of college attendance decision built on Heckman & Vytlacil (2005).

[17] The French higher education system includes the universities, which do not impose any entry selection, as well as the *Grandes Ecoles* and specialized technical colleges, which are selective.

streams of future log-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, (η_0, η_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 (ν_0, ν_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

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

where $G_k(X)$ denotes the consumption value associated with the schooling decision k.[18] After graduating from high school, individuals are supposed to make the decision which yields the highest expected utility. Thus, the selection equation corresponds exactly to Equation (2.1). 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 principle also account for the direct costs of post-secondary schooling, can be interpreted in this context as a truly non-pecuniary component, including the taste for schooling and preferences for future non-wage job attributes (as they may depend on higher education attendance).

5.2 The data

We use French data from the *Generation 1992* and *Generation 1998* surveys in order to estimate our schooling choice model.[19] 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 main advantage of these two databases is that they contain information on both educational and labor market histories (over the first five years following the exit from the educational system). The surveys also provide a set of individual covariates used as controls in our estimation procedure. As most of the individual covariates are observed in both dataset, we exploit the pooled dataset hereafter.

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

[19] Beffy et al. (2010) also rely on these data to estimate the influence of expected returns when choosing a college major.

Variable	Higher education attendees Mean	Std. dev.	High school level Mean	Std. dev.
Initial monthly log wage (1992 French Francs)	8.75	0.44	8.50	0.39
Secondary schooling track				
L (Humanities)	0.15	0.36	0.04	0.19
ES (Economics and Social Sciences)	0.17	0.38	0.04	0.19
S (Sciences)	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	0.02	0.16	0.02	0.15
Father born abroad	0.11	0.32	0.11	0.32
Mother born abroad	0.10	0.31	0.10	0.30
Entering the labor market in 1992	0.46	0.50	0.51	0.50
Entering the labor market in 1998	0.54	0.50	0.49	0.50
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
Other	0.06	0.24	0.06	0.24
Age in 6^{th} grade				
≤ 10	0.10	0.29	0.03	0.17
11	0.84	0.37	0.72	0.45
≥ 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	19,143		5,082	

Table 2: Descriptive statistics.

Our subsample of interest comprises 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 school 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. We drop individuals who only worked as temporary workers or were out of the labor force during the observation length, as their wages are not observed in the data. This finally leaves us with a large sample of 24,225 individuals. Although not common in the semiparametric literature estimating this kind of models, working with many observations is especially important for the semiparametric estimation procedure to perform well.[20] We report in Table 2 some descriptive statistics for the subsample of interest, according to higher education attendance. 79% of our sample (with a slight increase over the period, respectively 77.2% for *Generation 1992* and 80.6% for *Generation 1998*) attended higher education after graduating from high school. 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. We examine later on the sensitivity of our results to the inclusion of dropouts in the sample.

The functions $\psi_0(.), \psi_1(.)$ and $G(.)$ are assumed to depend on the secondary schooling track, whether the student is born abroad (and similarly for her parents), her year of entry into the labor market (1992 or 1998), her gender, her parental profession, her age in 6^{th} grade (i.e., her age of entry into junior high school)[21] and a dummy for living in Paris region (at the time of entry into junior high school). Given that following a vocational secondary schooling track seems to be a very strong predictor of higher education attendance (see Table 2), we also include in the set of regressors interactions between this variable and dummies for the year of entry into the labor market, gender and Paris region. Aside from this common set of regressors, we also include sector-specific variables, by supposing that the average local log-earnings of high school (resp. higher education) graduates affects $\psi_0(.)$ (resp. $\psi_1(.)$) alone. These variables, which are computed from the French Labor Force Surveys (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) for the high school and higher education graduates.[22] Migration costs imply 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.

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

[21]We use this variable as a proxy for ability since most of its variation stems from grade retention, which is quite common in France and mainly based on schooling performance. Students who neither repeat nor advance a grade before junior high school enter it at 11.

[22]More precisely, these variables were constructed by taking the average log-wages in the *departement* of residence at the time of entry into junior high school, weighted by the local rates of employment, over a 5-year time span centered respectively in 1992 or in 1998.

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, τ 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 year t, the 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.

As already mentioned, we do not observe incomes during the whole life cycle 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

$$y_{k,t} = \rho_k \mathbb{1}\{t - t_{0,k} + 1 \leq B\} + y_{k,t-1} + \nu_{k,t}, \tag{5.1}$$

where ρ_k denotes the alternative k-specific return to experience and $\nu_{k,t}$ is an alternative k-specific unobserved individual productivity term which is assumed to be independently and identically distributed over time, with mean zero. We introduce the dummy $\mathbb{1}\{t-t_{0,k}+1 \leq B\}$ to account for non significant marginal returns to experience after B years of work (see, e.g., Kuruscu, 2006, for a similar assumption on wage growth). We also suppose that $\nu_{k,t}$ is independent of D, so that ρ_k is simply identified by $\rho_k = E(y_{k,t} - y_{k,t-1}|D = k)$, for $t \leq B + t_{0,k} - 1$.

Now, let $\widetilde{\tau}_k = \tau^{t_{0,k}} \left(\frac{1-\tau^{A+1}}{1-\tau}\right)$, $C_k = \tau^{t_{0,k}} \left(\frac{\tau}{(1-\tau)^2}\right) \left(1 - \tau^B + B\tau^{A+1}(\tau - 1)\right)$ and

$$Y_k = \widetilde{\tau}_k y_{D,t_{0,D}} + \rho_k C_k.$$

Because $\widetilde{\tau}_D$, C_D and ρ_D are identified for given τ, A and B, we can identify $Y = DY_1 + (1-D)Y_0$. Moreover, under (5.1), we have $Y_k = E(Y_k^*|X, \eta_0, \eta_1, \nu_{k,t_{0,k}})$, which in turn 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 instead of the unobserved variable Y^*.

In practice, we set $\tau = 0.95$, $A = 45$ years, $B = 25$ years and estimate ρ_0 and ρ_1 to be respectively 0.025 and 0.042. These estimates 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-stage semiparametric procedure detailed in Section 3, with $\psi_k(X) = X'\beta_k$ and $G(X) = \delta_0 + X'\gamma_0$. Identification is secured here through the use of the average local log-earnings of high school and higher education graduates as sector-specific regressors. We use for the first step a mixture of probit (see, e.g., Coppejans, 2001) with $K_1 = 3$ mixture components.[23] The second step is performed with Newey (2009)'s series estimator, with $K_2 = 9$ approximating terms. We finally use for the last step the same specifications as in the Monte Carlo simulations.[24]

We also estimate bounds on the distribution of the *ex ante* treatment effect Δ, 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),

$$P(D = 0|X) = F_{\eta_\Delta}\left(\delta_0 + X'\alpha_0\zeta_0\right).$$

Therefore, we can obtain an estimator $\widehat{F}_{\eta_\Delta}(.)$ on $[\widehat{\underline{M}}, \widehat{\overline{M}}]$, the estimated support of $\delta_0 + X'\alpha_0\zeta_0$, by regressing nonparametrically $1 - D$ on the index $\widehat{\delta} + X'\widehat{\alpha}\widehat{\zeta}$. 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 (2.7) and (2.8). Bounds on the distribution of the *ex ante* treatment effect on the treated are estimated similarly, using (2.9). In practice, we consider a kernel estimator of F_{η_Δ} with a gaussian kernel, and a bandwidth $\tilde{h}_n = 1.6\sigma(\widehat{U})n^{-1/5}$.

5.4 Results

The first step estimates of $(\zeta, \beta_0, \beta_1)$ are displayed in Table 6 in Appendix C. Overall, the results for β_0 and β_1 display a quite similar pattern. In particular, the local average income variables that we use as sector-specific variables 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. However, some characteristics only affect the earnings of high school graduates or higher education attendees. This is in particular the case of gender, with high school male graduates earning significantly more than females. This is also the case of vocational secondary schooling

[23] 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.

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

tracks relative to technical tracks, which are positively related to earnings for high school graduates, while this is only true for male higher education attendees. Conversely, parental profession affects more significantly the earnings of higher education attendees than high school graduates, with negative signs associated with inactive, deceased or unemployed mother (referred to as "Other" in the Tables), relative to white collar professions. Similarly, higher education attendees whose mother is employed in an agricultural profession also earn significantly less.

The first column of Table 3 below reports the parameter estimates relative to the non-pecuniary component $G(.)$ which are obtained with the unconstrained specification, i.e. without assuming any restriction on the non-pecuniary component. The coefficients corresponding to the local average income of higher education and high school graduates are both not significant at the 10% level. This supports the idea that, as proxies for local labor market conditions, these variables have no reason to enter the non-pecuniary factors and should therefore only affect the probability of attendance through the *ex ante* returns. It also suggests that the data is consistent with a constrained specification where, e.g., the coefficient related to the local average income of high school graduates is set equal to zero.[25] As already shown in Section 4, our estimation procedure performs substantially better when using an exclusion restriction on $G(.)$. Hence, we focus on the constrained specification hereafter.

Several patterns emerge from the constrained estimates of $G(.)$ displayed in the second column of Table 3. First, as expected, the estimates are indeed substantially more precise than with the unconstrained specification. The results suggest that individuals attending a general secondary schooling track (namely L for Humanities, ES for Economics and Social Sciences and S for Sciences), relative to a technical track, value positively higher education attendance, with the related coefficients being significant at the 1% level.[26] 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 are in general tracks. The positive effect of entering the labor market in 1998 probably reflects the enlargement of access to

[25] We choose to impose the nullity of the coefficient associated with the local average income of high school graduates rather than the one of higher education graduates since (i) its point estimate in the unconstrained setting is much lower and (ii) the latter coefficient is close to the 10% significativity level.

[26] 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.

higher education which took place in France during the nineties. Individuals living in the Paris region also have a higher probability to attend higher education through these non-pecuniary factors, reflecting the large supply of post-secondary institutions in this area. 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 a parameter significant at the 1% level), possibly reflecting higher educational aspirations for males than for females (see, e.g., Page et al., 2007, for experimental evidence on this point). 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 junior 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. The positive effect on higher education attendance of living in the Paris region is significantly weaker for the individuals graduating from a vocational high school track. This result stresses once more the important explanatory power of the secondary schooling track. Consistent with the results of the unconstrained specification, the coefficient related to the local average income of higher education graduates is small, and here only significant at the 10% level. Finally, an estimation of the non-pecuniary component of each individual in the sample reveals that for 84% of them, this component is negative. Hence, we find, 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.

Variable	Unconstrained	Constrained
Constant (δ_0)	-0.185 (0.174)	-0.026 (0.155)
Local average income		
Higher education graduates	-0.026 (0.017)	-0.014* (0.008)
High school graduates	0.01 (0.012)	0
Secondary schooling track		
L	-0.288*** (0.087)	-0.142*** (0.054)
ES	-0.336*** (0.097)	-0.172*** (0.058)
S	-0.349*** (0.097)	-0.175*** (0.061)
Vocational	0.62** (0.248)	0.293* (0.164)
Technical	*Ref.*	*Ref.*
Born abroad	-0.084** (0.033)	-0.031 (0.021)
Father born abroad	-0.034* (0.02)	-0.005 (0.011)
Mother born abroad	0.003 (0.014)	-0.009 (0.013)
Entering the labor market in 1998 (relative to 1992)	-0.272*** (0.084)	-0.12** (0.051)
Male	-0.062*** (0.015)	-0.038*** (0.009)
Father's profession		
Farmer	-0.029 (0.02)	-0.023 (0.017)
Tradesman	-0.053*** (0.02)	-0.025** (0.011)
Executive	-0.105*** (0.034)	-0.054** (0.022)
Intermediate occupation	-0.071*** (0.025)	-0.035*** (0.011)
Blue collar	0.000 (0.012)	-0.004 (0.008)
Other	-0.036** (0.015)	-0.023** (0.011)
White collar	*Ref.*	*Ref.*
Mother's profession		
Farmer	0.091** (0.039)	0.057 (0.037)
Tradesman	0.021 (0.019)	-0.003 (0.011)
Executive	-0.056*** (0.02)	-0.023* (0.014)
Intermediate occupation	-0.018 (0.013)	-0.019* (0.011)
Blue collar	0.076*** (0.027)	0.019* (0.01)
Other	0.012 (0.014)	-0.01 (0.007)
White collar	*Ref.*	*Ref.*
Age in 6th grade		
≤ 10	-0.103*** (0.038)	-0.047** (0.024)
11	*Ref.*	*Ref.*
≥ 12	0.108*** (0.041)	0.056** (0.026)
Paris region	-0.082*** (0.025)	-0.03** (0.012)
Vocational × ...		
Entering the labor market in 1998	0.068** (0.029)	0.034 (0.024)
Male	-0.02 (0.021)	0.003 (0.014)
Paris region	0.126*** (0.048)	0.059** (0.029)

Standard errors, presented in parentheses, were computed by bootstrap with 200 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 below, respectively for the whole sample and for the subsample of higher education attendees. The streams of earnings were divided by 1,000 for scaling reasons, so that these returns must be compared to values which range from 0.7 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 $u \simeq 0.65$.[27] The upper bound of the distribution can be used to compute a lower bound $\underline{E}$ on the average return to higher education $E(Y_1 - Y_0)$.[28] We obtain $\underline{E} \simeq 0.12$, which is quite large since it is close to one standard deviation of Y. We also observe a large heterogeneity on these returns, with a range on the *ex ante* returns $E(Y_1 - Y_0|X, \eta_0, \eta_1)$ which is similar 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.

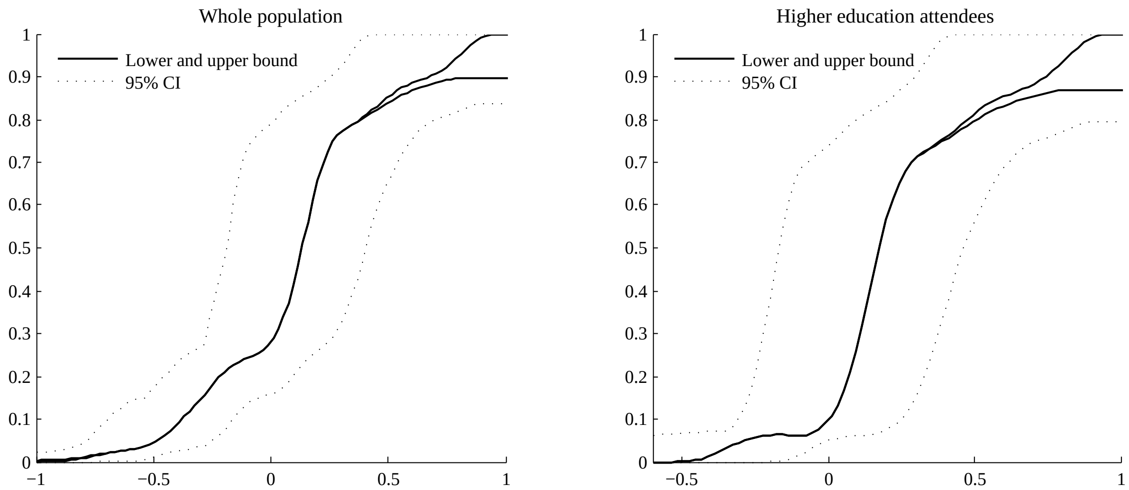

Figure 1: Distribution of the *ex ante* returns to higher education.

As expected, the distribution of the *ex ante* return 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 28% for the whole sample. Hence, about 10% of the

[27]Besides, the estimated cumulative distribution functions of the *ex ante* returns to higher education are increasing, which provides a check for the validity of our specification.

[28]Indeed, an integration by part shows that

$$E(Y_1 - Y_0) = \int_{-\infty}^{\infty} \left[\mathbb{1}\{u \geq 0\} - F_\Delta(u)\right] du.$$

This integral can be bounded below by the corresponding integrals on $\overline{F_\Delta}$. Note that we cannot obtain a finite upper bound on $E(Y_1 - Y_0)$ here because $\lim_{u \to +\infty} \widehat{\underline{F_\Delta}}(u) < 1$.

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, the probability of attending higher education would fall by 11.1 percentage points (from the predicted access rate, equal to 83.1%, to the probability of having a positive *ex ante* return, 72%) if non-pecuniary factors did not exist. For comparison purposes, this decrease in higher education attendance rate is eight times larger than, for instance, the 1.4 point decrease associated with a 10% permanent decrease in labor market earnings of higher education attendees.

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 (-0.326) is, in absolute terms, quantitatively much larger than the median *ex ante* return to higher education (0.133). Aside from their large median magnitude, non-pecuniary factors also have a fairly large dispersion, with an interquartile range equal to 0.239 which is nevertheless smaller than the interquartile range for *ex ante* returns (0.336).

Quartile	*Ex ante* return	Non-pecuniary factors
25%	-0.069	-0.430
50%	0.133	-0.326
75%	0.267	-0.191

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 show once more that non-pecuniary factors matter much when deciding whether to attend higher education. Indeed, the predicted attendance rate falls steeply, by more than 32 points, when making G vary from its first to its last decile. These estimates therefore suggest that the variation across individuals in non-pecuniary factors accounts for a very substantial part of the observed decisions to attend higher education. Overall, in line with recent evidence by Carneiro et al. (2003) and Beffy et al. (2010), non-pecuniary factors appear to be a key determinant of the decision to attend higher education.

Decile of G	Predicted attendance rate
10%, $G = -0.490$	0.952
90%, $G = 0.073$	0.630

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

5.5 Robustness checks

We address in the following two potential concerns about our results, namely the validity of our instrumental strategy and the robustness of our results to the assumptions made when computing the streams of earnings.[29]

5.5.1 Validity of the instrumental strategy

The validity of the results discussed above hinges on the exclusion restrictions between sectors. A reason why this identification strategy may not hold is that some individuals who attended higher education might actually face labor market conditions similar to the ones faced by those with a high school level. This might in particular be true for higher education dropouts, who enter the labor market without any post-secondary diploma (see, e.g., Kane & Rouse, 1995, for evidence from U.S. data of a small wage premium for some college, relative to high school graduation). In order to cope with this potential concern, we run our estimates without the $3,092$ higher education dropouts. By doing so, we focus on higher education graduation rather than attendance, in a similar spirit as in Carneiro et al. (2003). The resulting estimates of the non-pecuniary factors (see Panel 1, Table 7) are very similar to previously. Secondary schooling track, gender, father'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* return to higher education is also very similar to previously (see Figure 2) and remains within the confidence intervals of that of the baseline specification. Hence, the robustness of the results to the exclusion of higher education dropouts from the sample supports our exclusion restrictions.

One might also suspect that variations across *departements* in sector-specific average incomes could be correlated with geographical variations in sector-specific labor market productivity. If this were the case, the sector-specific regressors would be endogenous with

[29]Tables and figures that we refer to in this subsection are reported in Appendix C.

respect to the potential earnings, thus resulting in biased estimates. In order to deal with this issue, we include in the set of regressors the local proportion of individuals who graduated from high school with honours. This variable, which is computed from the *Panel 1989* dataset (French Ministry of Education), is used to control for differences across *departements* in productivity levels.[30] The estimates of the non-pecuniary factors (see Panel 2, Table 7) as well as of the distribution of the *ex ante* returns to education (see Figure 2) are robust to this alternative specification, suggesting that our estimates are likely not to be biased by the kind of mechanism discussed above.

5.5.2 Alternative computations of the streams of earnings

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 3 and 4 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. 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 seem overall robust to alternative computations of Y.[31]

6 Conclusion

This paper focuses on the effect of covariates on earnings and on the non-pecuniary component in a generalized Roy model. Our main theoretical contribution is to prove that the identification of the covariates effects entails the identification of the non-pecuniary component. The detailed structure of the model is indeed sufficient to recover this parameter. In particular, no exclusion restriction is required. Our approach does not hinge either on any assumption on the information set of the agents, as we do not impose any restriction on

[30] The *Panel 1989* is a longitudinal dataset that follows 22,000 students entering the 6th grade in 1989.

[31] We also estimate the streams of earnings where people are aware of their own annual increase ρ_i of log-earnings, instead of just anticipating an average increase. We estimate ρ_i by OLS and compute the corresponding streams of earnings. The signs of γ remain the same but no coefficient is significant anymore. This can be explained by i) the importance of the errors on the estimated ρ_i and ii) the fact that the sample we can use in this case comprises only 9,364 individuals.

the sector-specific productivity terms, apart from their independence with the covariates. Being agnostic on this issue is convenient since the determination of this information set is still an ongoing body of research (see, e.g., Cunha & Heckman, 2007).

We also contribute to the treatment effects literature by providing set identification results for the distribution of the *ex ante* treatment effects, in the absence of instrumental variables. We propose a three-stage 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 factors and expected returns to schooling in the decision to attend higher education. Consistent with the recent empirical evidence on this question, our main insight is that non-pecuniary factors are a key determinant of the attendance decision. From a policy point of view, our results suggest that a moderate increase in tuition fees, which is currently discussed to help finance the French higher education system, would only have a small detrimental effect on the higher education participation rate.

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 η_0 and η_1. From an economic point of view, providing identification results on this dependence is especially worthwhile 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 an alternative 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,

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

Now let us show that almost surely,

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

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

$$\eta_\Delta \geq \inf \mathcal{A}_x \Rightarrow \eta_\Delta \geq \psi_0(x) - \psi_1(x) + G(x).$$

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,

$$E(Y|D = 1, X) = \psi_1(\widetilde{X}_1) + h(P(D = 1|X)).$$

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

$$E(Y|D = 1, X) = \widetilde{\psi_1}(\widetilde{X}_1) + \widetilde{h}(P(D = 1|X)).$$

Then

$$(\widetilde{\psi_1} - \psi_1)(\widetilde{X}_1) + (\widetilde{h} - h)(P(D = 1|X)) = 0$$

By the measurably separation condition, this implies that $\widetilde{\psi_1}$ and ψ_1 are almost surely equal up to a constant. This constant is identified by Assumption 2.2. Thus, ψ_1 is identified. ψ_0 can be recovered by the same argument.

Theorem 2.2

The proof relies on Theorem 2.1 of D'Haultfoeuille & Maurel (2009). Their Assumptions 1 and 2 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,

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

This implies that Assumption 3 of D'Haultfoeuille & Maurel (2009) holds, and the result follows.

Theorem 2.3

First, note that

$$\frac{\partial q_0}{\partial x_1}(x_1, x_{-1}) = -\frac{\partial (T + G)}{\partial x_1}(x_1, x_{-1}) f_{\eta_\Delta} \left(T(x_1, x_{-1}) + G(x_1, x_{-1}) \right).$$

Thus, by Assumption 2.7, $\frac{\partial q_0}{\partial x_1}(x_1, x_{-1}) \neq 0$ as soon as $\frac{\partial (T+G)}{\partial x_1}(x_1, x_{-1}) \neq 0$. Hence, by Equation (2.6), $G(., x_{-1})$ is identified on the set $\mathcal{A}_{x_{-1}} = \{x_1 / \frac{\partial (T+G)}{\partial x_1}(x_1, x_{-1}) \neq 0\}$. If $\mathcal{A}_{x_{-1}}$ is confounded with the support of X_1 conditional on $X_{-1} = x_{-1}$, then $G(., .)$ is identified. Otherwise, let us consider $x_1 \notin \mathcal{A}_{x_{-1}}$. Because $\mathcal{A}_{x_{-1}} \neq \emptyset$ by assumption, either $\mathcal{A}_{x_{-1}} \cap (-\infty, x_1)$ or $\mathcal{A}_{x_{-1}} \cap (x_1, \infty)$ is nonempty. Suppose without loss of generality that the former set is nonempty, and let $\overline{x}_1$ denote its supremum. By definition, $\frac{\partial (T+G)}{\partial x_1} = 0$ on $(\overline{x}_1, x_1]$. Thus,

$$G(x_1, x_{-1}) = -T(x_1, x_{-1}) + G(\overline{x}_1, x_{-1}) + T(\overline{x}_1, x_{-1}).$$

Besides, by definition of the supremum, there exists a sequence $(x_{n1})_{n \in \mathbb{N}}$ which tends to $\overline{x}_1$ and such that $x_{n1} \in \mathcal{A}_{x_{-1}}$ for all n. As a result, it follows from the continuity of $(T + G)(., x_{-1})$ implied by Assumption 2.6 that $G(x_1, x_{-1})$ is identified by

$$G(x_1, x_{-1}) = -T(x_1, x_{-1}) + \lim_{n \to \infty} G(x_{n1}, x_{-1}) + T(x_{n1}, x_{-1}).$$

The result follows.

Theorem 3.1

Before establishing the result, let us introduce some notations. Let $f(.,\zeta)$ denote the density of $X'\zeta$, $q(u,\zeta) = E(D|X'\zeta = u)$, $r(.,\zeta) = q(.,\zeta) \times f(.,\zeta)$ and define $f_0(.) = f(.,\zeta_0)$, $q_0(.) = q(.,\zeta_0)$ and $r_0(.) = q_0(.)f_0(.)$. Consider the kernel estimators

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

and $\widehat{r}(.,\zeta) = \widehat{q}(.,\zeta) \times \widehat{f}(.,\zeta)$, where $\widehat{q}(.,\zeta)$ is defined by Equation (3.6). Let us also define $Z_i(\zeta) = \mathbb{1}\{X_i \in \mathcal{X}\}h(X_i'\zeta)$ and, for any $\mu = (r(.), f(.), \zeta, \widetilde{\beta}_0, \widetilde{\beta}_1)$,

$$V_i(\mu) = D_i X_i'\zeta - \int_{u_0}^{X_i'\zeta} \frac{r(u)}{f(u)} du.$$

We then let $W_i(\mu) = (1, D_i, V_i(\mu))'$. Thus, $\widehat{W_i} = W_i(\widehat{\mu})$ and $W_i = W_i(\mu_0)$, with $\widehat{\mu} = (\widehat{r}(.,\widehat{\zeta}), \widehat{f}(.,\widehat{\zeta}), \widehat{\zeta}, \widehat{\beta}_0, \widehat{\beta}_1)$ and $\mu_0 = (r_0, f_0, \zeta_0, \beta_0, \beta_1)$. Similarly, let

$$\varepsilon_i(\mu) = Y_i - X_i'\left(D_i\widetilde{\beta}_1 + (1 - D_i)\widetilde{\beta}_0\right).$$

Eventually, let $g(A_i,\theta,\mu) = Z_i(\zeta)(\varepsilon_i(\mu) - W_i(\mu)'\theta)$ and $g(A_i,\mu) = g(A_i,\theta_0,\mu)$, with $A_i = (D_i, Y_i, X_i)$. Then $E[g(A,\mu_0)] = 0$ and

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

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 ζ in the nonparametric estimator of q_0. The proof of the theorem proceeds in three steps.

Step 1. We first show that $\mu \mapsto \sum_{i=1}^n g(A_i,\mu)$ can be linearized in a convenient way. Recalling that $U_i = X_i'\zeta_0$, we let

$$\begin{aligned} G(A_i,\mu) &= \xi_i \frac{\partial Z_i}{\partial \zeta}(\zeta_0)'\zeta + Z_i(\zeta_0)\Bigg[- X_i'(D_i\widetilde{\beta}_1 + (1-D_i)\widetilde{\beta}_0) - \Bigg(D_i X_i'\zeta \\ & -q_0(U_i)X_i'\zeta - \int_{u_0}^{U_i} \frac{\partial q}{\partial \zeta}(u,\zeta_0)'\zeta + \frac{1}{f_0(u)}\left(r(u) - q_0(u)f(u)\right) du\Bigg)\alpha_0\Bigg]. \end{aligned}$$

Note that $\partial q/\partial\zeta(.,\zeta_0)$ exists under Assumptions 2.3 and 3.2, by Lemma 8.1. Let us also define $\widetilde{\mu} = (\widetilde{r}, \widetilde{f}, \widehat{\zeta}, \widehat{\beta}_0, \widehat{\beta}_1)$ where $\widetilde{r} = \widehat{r}(.,\zeta_0)$ and $\widetilde{f} = \widehat{f}(.,\zeta_0)$. We shall prove that

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

For that purpose, we use the decomposition

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

where, denoting by $h'(.)$ the vector of derivatives of $h(.)$ and $\widetilde{q}=\widetilde{r}/\widetilde{f}$, we let

$$\begin{aligned}
R_{1i} &= \xi_i \mathbb{1}\{X_i\in\mathcal{X}\}\left(h(\widehat{U}_i)-h(U_i)-(\widehat{U}_i-U_i)h'(U_i)\right),\\
R_{2i} &= \alpha_0 Z_i(\zeta_0)\left[\int_{U_i}^{\widehat{U}_i}\widehat{q}(u,\widehat{\zeta})du-q_0(U_i)(\widehat{U}_i-U_i)\right],\\
R_{3i} &= \alpha_0 Z_i(\zeta_0)\int_{u_0}^{U_i}\widehat{q}(u,\widehat{\zeta})-\widetilde{q}(u)-\frac{\partial q}{\partial\zeta}(u,\zeta_0)'(\widehat{\zeta}-\zeta_0)du,\\
R_{4i} &= \alpha_0 Z_i(\zeta_0)\int_{u_0}^{U_i}\widetilde{q}(u)-q_0(u)-\frac{1}{f_0(u)}\left(\widetilde{r}(u)-r_0(u)-q_0(u)(\widetilde{f}(u)-f_0(u))\right)du,\\
R_{5i} &= \left[\varepsilon_i(\widehat{\mu})-\varepsilon_i(\mu_0)-(W_i(\widehat{\mu})-W_i(\mu_0))'\theta_0\right]\left[Z_i(\widehat{\zeta})-Z_i(\zeta_0)\right].
\end{aligned}$$

We now check that for all $k\in\{1,...,5\}$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ki}=o_P(1)$.

$-R_{1i}$: by Assumption 3.2, there exists C_0 such that $\|X\|\leq C_0$, where $\|.\|$ denotes the euclidian norm. Then, by the Cauchy-Schwarz inequality, $|\widehat{U}_i-U_i|\leq C_0||\widehat{\zeta}-\zeta_0||$. Thus, by Assumptions 3.4 and 3.7,

$$\begin{aligned}
\sqrt{n}\max_{i=1,...,n}\left|h(\widehat{U}_i)-h(U_i)-(\widehat{U}_i-U_i)h'(U_i)\right| &\leq \sqrt{n}M\max_{i=1,...,n}|\widehat{U}_i-U_i|^2\\
&\leq MC_0^2\sqrt{n}||\widehat{\zeta}-\zeta_0||^2\\
&= o_P(1),
\end{aligned}$$

where $M=\|\max|h''|\|$. 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 $\mathcal{S}_0=\{x'\zeta_0,x\in\mathcal{X}\}$. By definition, $\mathcal{S}_0\subsetneq\mathcal{S}$, where $\mathcal{S}$ denotes the support of U. Besides, by definition, $Z_i(\zeta_0)=Z_i(\zeta_0)\mathbb{1}\{U_i\in\mathcal{S}_0\}$. Moreover, for all i such that $\widehat{U}_i\in\mathcal{S}_0$, there exists, by the mean value theorem, $\widetilde{U}_i=tU_i+(1-t)\widehat{U}_i$, with $t\in[0,1]$, such that $\int_{U_i}^{\widehat{U}_i}q_0(u)du=q_0(\widetilde{U}_i)(\widehat{U}_i-U_i)$. Thus, when $\widehat{U}_i\in\mathcal{S}_0$,

$$\begin{aligned}
\|R_{2i}\| &= \left\|\alpha_0 Z_i(\zeta_0)\mathbb{1}\{U_i\in\mathcal{S}_0\}\left\{\int_{U_i}^{\widehat{U}_i}\left[\widehat{q}(u,\widehat{\zeta})-q_0(u)\right]du+\int_{U_i}^{\widehat{U}_i}q_0(u)du-q_0(U_i)(\widehat{U}_i-U_i)\right\}\right\|\\
&\leq C_1\left|\widehat{U}_i-U_i\right|\left[\sup_{u\in\mathcal{S}_0}\left|\widehat{q}(u,\widehat{\zeta})-q_0(u)\right|+\max_{i:\widehat{U}_i\in\mathcal{S}}\left|q_0(\widetilde{U}_i)-q_0(U_i)\right|\right]\\
&\leq C_0C_1\left\|\widehat{\zeta}-\zeta_0\right\|\left[\sup_{u\in\mathcal{S}_0}\left|\widehat{q}(u,\widehat{\zeta})-q_0(u)\right|+\max_{i:\widehat{U}_i\in\mathcal{S}}\left|q_0(\widetilde{U}_i)-q_0(U_i)\right|\right],
\end{aligned}$$

where $C_1 > 0$ is a constant such that $\|\alpha_0 Z_i(\zeta_0)\| \leq C_1$, which exists by Assumptions 3.2 and 3.7. Besides, because $\widehat{q}(., \widehat{\zeta})$ and $q_0(.)$ are bounded above by 1, we have, when $\widehat{U}_i \notin \mathcal{S}_0$,

$$\|R_{2i}\| \leq 2C_0C_1 \left\|\widehat{\zeta} - \zeta_0\right\| \mathbb{1}\{U_i \in \mathcal{S}_0\}.$$

Hence,

$$\begin{aligned}\|R_{2i}\| \ \leq \ & C_0C_1 \left\|\widehat{\zeta} - \zeta_0\right\| \left[\sup_{u \in \mathcal{S}_0} \left|\widehat{q}(u, \widehat{\zeta}) - q_0(u)\right| + \max_{i:\widehat{U}_i \in \mathcal{S}} \left|q_0(\widetilde{U}_i) - q_0(U_i)\right| \right. \\ & \left. +2\mathbb{1}\{U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\}\right]. \qquad (7.4)\end{aligned}$$

By Assumption 3.4, $\sqrt{n}\left\|\widehat{\zeta} - \zeta_0\right\| = O_P(1)$. Let us now show that the term into brackets in (7.4) is a $o_P(1)$. By Lemma 8.2, $\sup_{u\in\mathcal{S}_0} |\widehat{q}(u, \widehat{\zeta}) - q_0(u)| = o_P(1)$. Now fix $\varepsilon > 0$. Because $q_0(.)$ is continuous by Assumption 2.3 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,

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

Because $|\widetilde{U}_i - U_i| \leq |\widehat{U}_i - U_i| \leq C_0\left\|\widehat{\zeta} - \zeta_0\right\|$, the right-hand side tends to one. This proves that

$$\max_{i:\widehat{U}_i \in \mathcal{S}} \left|q_0(\widetilde{U}_i) - q_0(U_i)\right| = o_P(1).$$

It remains to show that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\} = o_P(1). \qquad (7.5)$$

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 $P(U_i \in \mathcal{S}_K) < \varepsilon/2$. For n large enough, $P(C_0\left\|\widehat{\zeta} - \zeta_0\right\| > K) < \varepsilon/2$. Because $|U_i - \widehat{U}_i| \leq C_0\left\|\widehat{\zeta} - \zeta_0\right\|$, we have, for n large enough,

$$\begin{aligned}P\left(U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\right) \ & \leq \ \frac{\varepsilon}{2} + P\left(U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0, C_0||\widehat{\zeta} - \zeta_0|| \leq K\right) \\ & \leq \ \frac{\varepsilon}{2} + P(U_i \in \mathcal{S}_K) \\ & \leq \ \varepsilon.\end{aligned}$$

Because ε was arbitrary, this proves that

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

This establishes (7.5) 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 $\widetilde{\zeta}_u$ in the segment between ζ_0 and $\widehat{\zeta}$ such that

$$\widehat{q}(u,\widehat{\zeta}) - \widetilde{q}(u) = \frac{\partial \widehat{q}}{\partial \zeta}(u,\widetilde{\zeta}_u)'(\widehat{\zeta}-\zeta_0).$$

Because U_i is bounded, there exists C_2 such that $|U_i - u_0| < C_2$. Thus,

$$\begin{aligned}
|R_{3i}| &= \|\alpha_0 Z_i(\zeta_0)\| \left| \left[\int_{u_0}^{U_i} \frac{\partial \widehat{q}}{\partial \zeta}(u,\widetilde{\zeta}_u) - \frac{\partial q}{\partial \zeta}(u,\zeta_0) du\right]' (\widehat{\zeta}-\zeta_0)\right| \mathbb{1}\{U_i \in \mathcal{S}_0\} \\
&\leq C_1 C_2 \left\|\widehat{\zeta}-\zeta_0\right\| \sup_{u\in\mathcal{S}_0} \left\| \frac{\partial \widehat{q}}{\partial \zeta}(u,\widetilde{\zeta}_u) - \frac{\partial q}{\partial \zeta}(u,\zeta_0)\right\|.
\end{aligned}$$

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

$$\begin{aligned}
|R_{4i}| &\leq C_1 \mathbb{1}\{U_i \in \mathcal{S}_0\} \int_{u_0}^{U_i} \frac{1}{\widetilde{f}(u) f_0(u)} \left[1 + |q_0(u)|\right] \left[|\widetilde{f}(u) - f_0(u)|^2 + |\widetilde{r}(u) - r_0(u)|^2\right] du \\
&\leq \frac{2C_1C_2}{\inf_{u\in\mathcal{S}_0} \widetilde{f}(u) \inf_{u\in\mathcal{S}_0} f_0(u)} \left[\left(\sup_{u\in\mathcal{S}_0} |\widetilde{f}(u) - f_0(u)|\right)^2 + \left(\sup_{u\in\mathcal{S}_0} |\widetilde{r}(u) - r_0(u)|\right)^2\right] \quad (7.6)
\end{aligned}$$

Assumption 3.2 implies that the density of U_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 in the right-hand side of (7.6) 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

$$\begin{aligned}
&\left|\varepsilon_i(\widehat{\mu}) - \varepsilon_i(\mu_0) - (W_i(\widehat{\mu}) - W_i(\mu_0))'\theta_0\right| \mathbb{1}\{X_i \in \mathcal{X}\} \\
= &\left| X_i'(D_i(\beta_1 - \widehat{\beta}_1) + (1-D_i)(\beta_0 - \widehat{\beta}_0)) + \left(D_i(U_i - \widehat{U}_i) + \int_{U_i}^{\widehat{U}_i} \widehat{q}(u,\widehat{\zeta}) du \right.\right. \\
&\left.\left. + \int_{u_0}^{U_i} \left[\widehat{q}(u,\widehat{\zeta}) - q_0(u)\right] du\right)\alpha_0\right| \mathbb{1}\{X_i \in \mathcal{X}\} \\
\leq & C_0\left(\left\|\widehat{\beta}_1 - \beta_1\right\| + \left\|\widehat{\beta}_0 - \beta_0\right\| + 2|\alpha_0| \left\|\widehat{\zeta} - \zeta_0\right\|\right) + C_2|\alpha_0| \sup_{u\in\mathcal{S}_0} |\widehat{q}(u,\widehat{\zeta}) - q_0(u)|.
\end{aligned}$$

where the first term of the upper bound follows from the Cauchy-Schwarz inequality.

Besides, with probability approaching one, there exists a compact which contains $\widehat{U}_i$ and U_i for all i. Thus, because h' is continuous, there exists $C_3 > 0$ such that, with probability approaching one,

$$\left\| Z_i(\widehat{\zeta}) - Z_i(\zeta_0) \right\| \leq C_3 \left\| \widehat{\zeta} - \zeta_0 \right\|.$$

Hence, with probability approaching one,

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{5i} \right| &\leq \left[C_0 C_3 \sqrt{n} \left\| \widehat{\zeta} - \zeta_0 \right\| \right] \left[\left\| \widehat{\beta}_1 - \beta_1 \right\| + \left\| \widehat{\beta}_0 - \beta_0 \right\| + 2|\alpha_0| \left\| \widehat{\zeta} - \zeta_0 \right\| \right. \\ & \left. + C_2 |\alpha_0| \sup_{u \in \mathcal{S}_0} |\widehat{q}(u, \widehat{\zeta}) - q_0(u)| \right]. \end{aligned}$$

By Assumption 3.4, 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 $1/\sqrt{n}\sum_{i=1}^n G(A_i, \widetilde{\mu} - \mu_0)$ can be linearized. Let $\kappa_0 = (\zeta_0, \beta_1, \beta_0)'$ and $\widehat{\kappa} = (\widehat{\zeta}, \widehat{\beta}_1, \widehat{\beta}_0)'$. We have

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

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

$$\begin{aligned} P_{1i} &= \xi_i \frac{\partial Z_i}{\partial \zeta}(\zeta_0)' - \alpha_0 Z_i(\zeta_0) \left(D_i X_i' - q_0(U_i) X_i' - \int_{u_0}^{U_i} \frac{\partial q}{\partial \zeta'}(u, \zeta_0) du \right) \\ P_{2i} &= -Z_i(\zeta_0) D_i X_i' \\ P_{3i} &= -Z_i(\zeta_0)(1 - D_i) X_i' \\ \widetilde{G}(A_i, \widetilde{r}, \widetilde{f}) &= \alpha_0 Z_i(\zeta_0) \int_{u_0}^{U_i} (1/f_0(u))(\widetilde{r}(u) - q_0(u)\widetilde{f}(u)) du. \end{aligned}$$

By the weak law of large numbers,

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

Moreover, by Assumptions 3.4 and 3.5,

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

Thus,

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

where

$$\Omega_{1i} = E[P]' \left(\chi_i, \chi_{1i}, \chi_{0i} \right)'. \tag{7.8}$$

Thus, it suffices to focus on the nonparametric part of G, $\widetilde{G}(A_i, \widetilde{r}, \widetilde{f})$. The main insight here is that $\widetilde{G}$ is nearly the linearized part of the consumer surplus example of Newey &

McFadden (1994, p. 2204), except that their b is replaced by U_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

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

where $\Omega_{2i} = \alpha_0 Z_i(\zeta_0)(1-F_0(U_i))\mathbb{1}\{U_i \geq u_0\}(D_i - q_0(U_i))/f_0(U_i)$, $F_0(.)$ denoting the cumulative distribution function of U. 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,

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

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

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

Now,

$$Z_i(\widehat{\zeta})W_i(\widehat{\mu})' = Z_i(\zeta_0)W_i(\mu_0)' + Z_i(\widehat{\zeta})(W_i(\widehat{\mu})-W_i(\mu_0))' + (Z_i(\widehat{\zeta})-Z_i(\zeta_0))W_i(\mu_0)'.$$

Besides, by Assumption 3.7, $\left\|Z_i(\widehat{\zeta})-Z_i(\zeta_0)\right\| \leq C_3\left\|\widehat{\zeta}-\zeta_0\right\|$ for a given $C_3>0$. Moreover, reasoning as with R_{5i}, we get

$$\|W_i(\widehat{\mu})-W_i(\mu_0)\| \leq 2C_0\left\|\widehat{\zeta}-\zeta_0\right\| + C_2\sup_{u\in\mathcal{S}_0}|\widehat{q}(u,\widehat{\zeta})-q_0(u)|.$$

Finally, $\|W_i(\mu_0)\|$ and $\left\|Z_i(\widehat{\zeta})\right\|$ are bounded with probability approaching one. As a result,

$$\frac{1}{n}\sum_{i=1}^{n} Z_i(\widehat{\zeta})W_i(\widehat{\mu})' = \frac{1}{n}\sum_{i=1}^{n} Z_i(\zeta_0)W_i(\mu_0)' + o_P(1).$$

Thus, by the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} Z_i(\widehat{\zeta})W_i(\widehat{\mu})' \xrightarrow{P} E(Z(\zeta_0)W(\mu_0)') = E(ZW').$$

Eventually, by Slutski's lemma, and given that $g(A,\mu_0)=Z\xi$,

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

This concludes the proof.

8 Appendix B: technical lemmas

Lemma 8.1 *Suppose that Assumptions 2.3 and 3.2 hold. Then, for all $u \in \mathcal{S}$, the support of U, $\zeta \mapsto f(u,\zeta)$ and $\zeta \mapsto r(u,\zeta)$, the density of $X'\zeta$ and the derivative of $u \mapsto E(D\mathbb{1}\{X'\zeta \leq u\})$ respectively, admit partial derivatives at ζ_0 which satisfy:*

$$\frac{\partial f}{\partial \zeta}(u,\zeta_0) = -\left(E\left[X|U=u\right] f_0(u)\right)' \tag{8.1}$$

$$\frac{\partial r}{\partial \zeta}(u,\zeta_0) = -\left(E\left[DX|U=u\right] f_0(u)\right)' \tag{8.2}$$

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 δ_k denote the vector of dimension p, with 1 at the k-th component and 0 elsewhere. We have

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

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

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

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

$$\frac{\partial F}{\partial \zeta_k}(u,\zeta_0) = -E\left[X_k|U=u\right] f_0(u).$$

By Assumption 3.2 once more, $u \mapsto \partial F/\partial\zeta_k(u,\zeta_0)$ is continuously differentiable and

$$\frac{\partial^2 F}{\partial u \partial \zeta}(u,\zeta_0) = -\left(E\left[X|U=u\right] f_0(u)\right)'.$$

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

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

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

Then differentiability of $u \mapsto \partial G_0/\partial\zeta(u,\zeta)$ 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 closed interval $\mathcal{S}'$ strictly included in the interior of $\mathcal{S}$ and for all $\zeta_{u,n}$ such that $\sup_{u\in\mathcal{S}'}\|\zeta_{u,n}-\zeta_0\| = O_P(1/\sqrt{n})$, we have,*

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

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_{u,n}) - \frac{\partial q}{\partial\zeta}(u,\zeta_0)\right\| \;=\; o_P(1) \tag{8.4}$$

Proof: we first write

$$\sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - q_0(u)| \le \sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - \widehat{q}(u,\zeta_0)| + \sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_0) - q_0(u)| \tag{8.5}$$

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

$$\begin{aligned}
\sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - \widehat{q}(u,\zeta_0)| &= \sup_{u\in\mathcal{S}'}\frac{\left|(\widehat{r}(u,\zeta_{u,n}) - \widehat{r}(u,\zeta_0)) + \widehat{q}(u,\zeta_{u,n})(\widehat{f}(u,\zeta_0) - \widehat{f}(u,\zeta_{u,n}))\right|}{\widehat{f}(u,\zeta_0)} \\
&\le \sup_{u\in\mathcal{S}'}\frac{1}{\widehat{f}(u,\zeta_0)}\left[|\widehat{r}(u,\zeta_{u,n}) - \widehat{r}(u,\zeta_0)| + \left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right|\right] \\
&\le \frac{1}{\inf_{u\in\mathcal{S}'}\widehat{f}(u,\zeta_0)}\left[\sup_{u\in\mathcal{S}'}|\widehat{r}(u,\zeta_{u,n}) - \widehat{r}(u,\zeta_0)|\right. \\
&\qquad\qquad \left. + \sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right|\right].
\end{aligned} \tag{8.6}$$

Let us prove that

$$\sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right| = o_P(1) \tag{8.7}$$

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

$$\begin{aligned}
\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right| &\le \frac{1}{nh_n}\sum_{i=1}^n\left|K\left(\frac{u - X_i'\zeta_{u,n}}{h_n}\right) - K\left(\frac{u - X_i'\zeta_0}{h_n}\right)\right| \\
&\le \frac{C_4C_0\,\|\zeta_{u,n}-\zeta_0\|}{h_n^2} \\
&\le \frac{C_4C_0\sup_{u\in\mathcal{S}'}\|\zeta_{u,n}-\zeta_0\|}{h_n^2} = O_p\left(\frac{1}{\sqrt{n}h_n^2}\right).
\end{aligned}$$

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

$$\inf_{u\in\mathcal{S}'}\widehat{f}(u,\zeta_0) \ge -\sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right| + \inf_{u\in\mathcal{S}'}f_0(u),$$

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

$$\frac{1}{\inf_{u\in\mathcal{S}'}\widehat{f}(u,\zeta_0)} = O_p(1).$$

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 $\sup_{u\in\mathcal{S}'}|\widehat{f}(u,\zeta_0) - f_0(u)| = o_P(1)$ and similarly for $\widehat{r}(.,\zeta_0)$. This establishes (8.3).

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

$$\sup_{u\in\mathcal{S}'}\left|\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) - \frac{\partial q}{\partial\zeta}(u,\zeta_0)\right| = o_P(1) \tag{8.8}$$

We have

$$\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) = \frac{1}{\widehat{f}(u,\zeta_0)}\left[\frac{\partial\widehat{r}}{\partial\zeta}(u,\zeta_0) - \widehat{q}(u,\zeta_0)\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0)\right].$$

and similarly for $\partial q/\partial\zeta(u,\zeta_0)$. Thus,

$$\begin{aligned}
&\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) - \frac{\partial q}{\partial\zeta}(u,\zeta_0)\\
= \;& \frac{1}{\widehat{f}(u,\zeta_0)}\left\{\left[\frac{\partial\widehat{r}}{\partial\zeta}(u,\zeta_0) - \frac{\partial r}{\partial\zeta}(u,\zeta_0)\right] - \frac{\partial r}{\partial\zeta}(u,\zeta_0)\left[\frac{\widehat{f}(u,\zeta_0) - f_0(u)}{f_0(u)}\right]\right\}\\
&-\frac{\widehat{q}(u,\zeta_0)}{\widehat{f}(u,\zeta_0)}\left[\left(\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0) - \frac{\partial f}{\partial\zeta}(u,\zeta_0)\right) - \frac{\partial f/\partial\zeta(u,\zeta_0)}{f_0(u)}\left(\widehat{f}(u,\zeta_0) - f_0(u)\right)\right]\\
&-\frac{\partial f/\partial\zeta(u,\zeta_0)}{f_0(u)}\left(\widehat{q}(u,\zeta_0) - q_0(u)\right).
\end{aligned}$$

By what precedes, $\inf_{u\in\mathcal{S}'}\widehat{f}(u,\zeta_0)$ tends in probability to $\inf_{u\in\mathcal{S}'} f_0(u) > 0$, while $\sup_{u\in\mathcal{S}'}|\widehat{f}(u,\zeta_0) - f_0(u)| = o_P(1)$. Besides, $\widehat{q}(.,\zeta_0)$ is bounded by 1 and by Lemma 8.1, $\partial f/\partial\zeta(.,\zeta_0)$ is continuous on the compact set $\mathcal{S}$ and thus is bounded on this set. Thus, it suffices to prove that

$$\sup_{u\in\mathcal{S}'}\left|\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0) - \frac{\partial f}{\partial\zeta}(u,\zeta_0)\right| = o_P(1) \tag{8.9}$$

and similarly for r_0. By Lemma 8.1, $u \mapsto \partial f/\partial\zeta(u,\zeta_0)$ is the derivative of $-E(X|U = u)f_0(u)$. 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

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_{u,n})-\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0)\right\|=o_P(1)$$

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

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_{u,n})-\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0)\right\|=o_P(1) \tag{8.10}$$

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,

$$\begin{aligned}\left\|\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_{u,n})-\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0)\right\| &\leq \frac{1}{nh_n^2}\sum_{i=1}^{n}\|X_i\|\left|K'\left(\frac{u-X_i'\zeta_{u,n}}{h_n}\right)-K'\left(\frac{u-X_i'\zeta_0}{h_n}\right)\right| \\ &\leq \frac{C_5C_0^2\,\|\zeta_{u,n}-\zeta_0\|}{h_n^3}=O_p\left(\frac{1}{\sqrt{n}h_n^3}\right).\end{aligned}$$

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	ζ	β_0	β_1
Local average income			
Higher education graduates	1.541*** (0.087)	0	0.019*** (0.004)
High school graduates	-1 (0)	0.022*** (0.004)	0
Secondary schooling track			
L	9.348*** (0.452)	-0.07* (0.039)	-0.011 (0.025)
ES	9.899*** (0.416)	-0.043 (0.04)	-0.002 (0.027)
S	10.133*** (0.426)	-0.055 (0.042)	-0.012 (0.026)
Vocational	-29.131*** (0.488)	0.247** (0.106)	-0.086 (0.094)
Technical	*Ref.*	*Ref.*	*Ref.*
Born abroad	1.727*** (0.46)	-0.006 (0.017)	0.000 (0.010)
Father born abroad	1.26** (0.451)	-0.011 (0.009)	0.011* (0.006)
Mother born abroad	1.591*** (0.464)	-0.018* (0.011)	0.007 (0.007)
Entering the labor market in 1998 (relative to 1992)	9.133*** (0.447)	0.097*** (0.035)	0.173*** (0.024)
Male	-0.298 (0.401)	0.043*** (0.008)	-0.001 (0.003)
Father's profession			
Farmer	2.291*** (0.434)	-0.012 (0.012)	0.014 (0.009)
Tradesman	1.289*** (0.43)	-0.008 (0.009)	-0.005 (0.005)
Executive	3.897*** (0.422)	-0.025 (0.016)	0.005 (0.011)
Intermediate occupation	1.799*** (0.457)	0.000 (0.009)	0.004 (0.007)
Blue collar	-0.49 (0.418)	0.008 (0.006)	-0.007 (0.004)
Other	1.309*** (0.432)	-0.013 (0.009)	-0.008 (0.006)
White collar	*Ref.*	*Ref.*	*Ref.*
Mother's profession			
Farmer	-6.343*** (0.51)	0.042* (0.025)	-0.038** (0.018)
Tradesman	-0.328 (0.488)	0.008 (0.01)	-0.002 (0.006)
Executive	1.279*** (0.469)	-0.01 (0.012)	-0.006 (0.006)
Intermediate occupation	0.899* (0.489)	-0.001 (0.009)	-0.001 (0.006)
Blue collar	-1.075** (0.438)	0.006 (0.008)	0.002 (0.006)
Other	-0.315 (0.411)	-0.003 (0.006)	-0.019*** (0.004)
White collar	*Ref.*	*Ref.*	*Ref.*
Age in 6th grade			
≤ 10	3.825*** (0.465)	-0.028 (0.017)	0.007 (0.01)
11	*Ref.*	*Ref.*	*Ref.*
≥ 12	-5.07*** (0.425)	0.035* (0.018)	-0.019 (0.013)
Paris region	1.181*** (0.453)	0.003 (0.012)	-0.002 (0.004)
Vocational $\times$ ...			
Entering the labor market in 1998	-1.012** (0.499)	-0.033* (0.018)	-0.021 (0.015)
Male	1.622*** (0.477)	-0.016 (0.01)	0.022** (0.01)
Paris region	-4.402*** (0.521)	0.023 (0.022)	-0.013 (0.018)

Standard errors, presented in parentheses, were computed by bootstrap with 200 bootstrap sample replicates. Significativity levels: *** (1%), ** (5%) and * (10%).

Table 6: First step estimates.

Variable	Panel 1	Panel 2	Panel 3	Panel 4
Constant (δ_0)	-0.016 (0.171)	0.006 (0.175)	-0.028 (0.164)	-0.024 (0.155)
Local average income				
Higher education graduates	-0.01 (0.007)	-0.013* (0.008)	-0.01 (0.008)	-0.014* (0.008)
Local rate of honours		-0.014 (0.031)		
Secondary schooling track				
L	-0.128*** (0.046)	-0.132***(0.049)	-0.117** (0.059)	-0.142*** (0.054)
ES	-0.154*** (0.05)	-0.162*** (0.052)	-0.15** (0.063)	-0.172*** (0.058)
S	-0.146*** (0.051)	-0.164*** (0.054)	-0.135** (0.066)	-0.175*** (0.061)
Vocational	0.227 (0.226)	0.351** (0.173)	0.251 (0.175)	0.293* (0.165)
Technical	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Born abroad	-0.02 (0.02)	-0.032 (0.02)	-0.03 (0.022)	-0.032 (0.021)
Father born abroad	0 (0.01)	-0.005 (0.011)	-0.006 (0.012)	-0.005 (0.011)
Mother born abroad	-0.011 (0.011)	-0.006 (0.012)	-0.009 (0.014)	-0.009 (0.013)
Entering the labor market in 1998 (relative to 1992)	-0.094*** (0.034)	-0.106** (0.045)	-0.113** (0.055)	-0.12** (0.051)
Male	-0.061*** (0.012)	-0.043*** (0.008)	-0.044*** (0.009)	-0.038*** (0.009)
Father's profession				
Farmer	-0.02 (0.016)	-0.022 (0.016)	-0.018 (0.018)	-0.023 (0.017)
Tradesman	-0.021** (0.009)	-0.026** (0.013)	-0.02* (0.012)	-0.025** (0.011)
Executive	-0.051** (0.023)	-0.053** (0.022)	-0.043* (0.024)	-0.055** (0.022)
Intermediate occupation	-0.034** (0.014)	-0.04*** (0.015)	-0.03** (0.013)	-0.035*** (0.012)
Blue collar	-0.009 (0.007)	-0.007 (0.007)	-0.005 (0.009)	-0.004 (0.008)
Other	-0.016 (0.011)	-0.018 (0.011)	-0.021* (0.012)	-0.023** (0.011)
White collar	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Mother's profession				
Farmer	0.049 (0.034)	0.045 (0.03)	0.049 (0.039)	0.057 (0.037)
Tradesman	-0.008 (0.01)	0.002 (0.012)	-0.006 (0.012)	-0.003 (0.011)
Executive	-0.017 (0.012)	-0.018 (0.012)	-0.017 (0.015)	-0.023* (0.014)
Intermediate occupation	-0.018 (0.012)	-0.017 (0.011)	-0.019 (0.012)	-0.019* (0.011)
Blue collar	0.011 (0.007)	0.017* (0.01)	0.016 (0.01)	0.019* (0.01)
Other	-0.008 (0.007)	-0.01 (0.006)	-0.009 (0.007)	-0.01 (0.007)
White collar	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Age in 6th grade				
≤ 10	-0.037* (0.021)	-0.039** (0.019)	-0.033 (0.025)	-0.047** (0.024)
11	*Ref.*	*Ref.*	*Ref.*	*Ref.*
≥ 12	0.06* (0.036)	0.05** (0.024)	0.048* (0.028)	0.056** (0.026)
Paris region	-0.024* (0.012)	-0.023* (0.012)	-0.018 (0.014)	-0.03** (0.012)
Vocational × ...				
Entering the labor market in 1998	0.019 (0.023)	0.04* (0.023)	0.02 (0.026)	0.034 (0.024)
Male	0.019 (0.016)	0.004 (0.015)	0.008 (0.015)	0.003 (0.014)
Paris region	0.045* (0.025)	0.052* (0.028)	0.038 (0.032)	0.059** (0.029)

In Panel 1, the higher education dropouts are excluded from the sample. In Panel 2, the local rate of honours is included in the estimation. In Panel 3 and 4, 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 200 sample replicates. Significativity levels: *** (1%), ** (5%) and * (10%).

Table 7: Estimates of non-pecuniary factors: robustness checks.

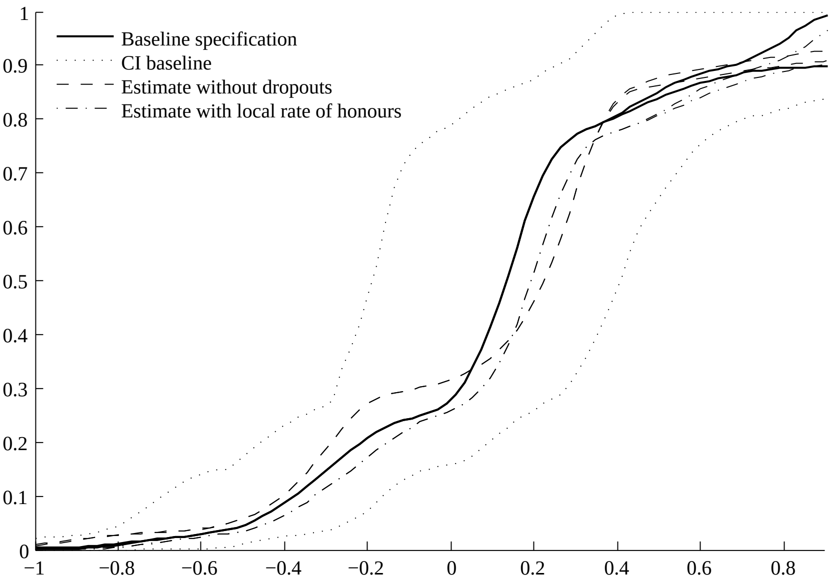

Figure 2: *Ex ante* returns to higher education: robustness of the instrumental strategy.

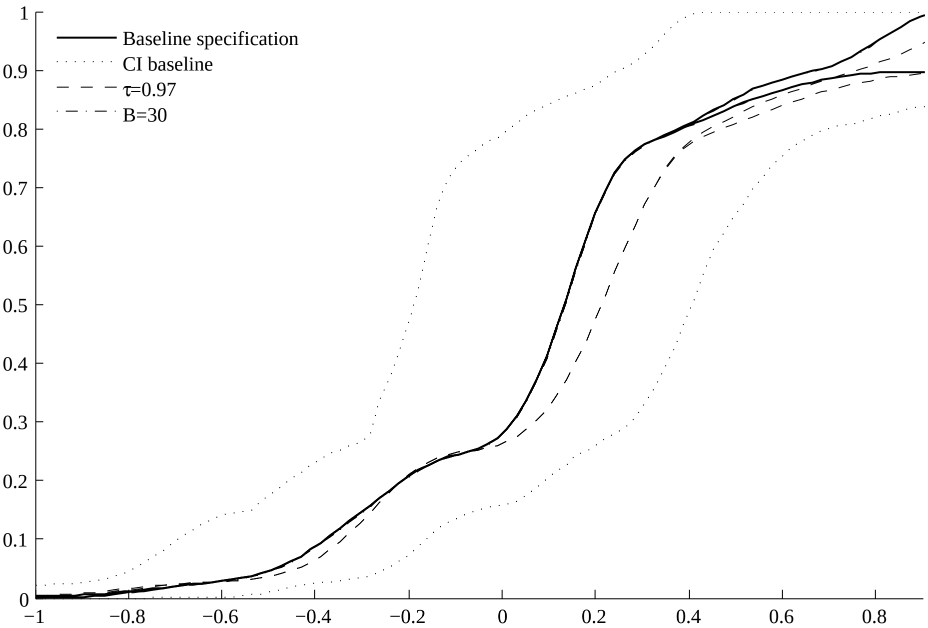

Figure 3: *Ex ante* returns to higher education under alternative computations of the streams of earnings.

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. (2010), 'Nonparametric identification and estimation in a roy model with common non-pecuniary returns', *Forthcoming in the Journal of Business and Economic Statistics* .

Beffy, M., Fougère, D. & Maurel, A. (2010), 'Choosing the field of study in post-secondary education: Do expected earnings matter?', *Forthcoming in the Review of Economics and Statistics* .

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., Heckman, J. & Vytlacil, E. (2010), 'Estimating marginal returns to education', *Forthcoming in the American Economic Review* .

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.

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

Heckman, J. J. & Honoré, 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. J. & Sedlacek, G. (1985), 'Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market', *Journal of Political Economy* **93**, 1077–1125.

Heckman, J. J. & Sedlacek, G. (1990), 'Self-selection and the distribution of hourly wages', *Journal of Labor Economics* **8**, S329–S363.

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.

Kane, T. & Rouse, C. E. (1995), 'Labor-market returns to two- and four-year college', *The American Econonomic Review* **85**, 600–614.

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.

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.

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.

Page, L., Garboua, L. & Montmarquette, C. (2007), 'Aspiration levels and educational choices: An experimental study', *Economics of Education Review* **26**, 747–757.

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.