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

Xavier D’Haultfœuille † Arnaud Maurel‡

First version: December 2009
This version: May 2011

Abstract

This paper considers the identification and estimation of an extension of Roy’s model (1951) of sectoral choice, which includes a non-pecuniary component in the selection equation and allows for uncertainty on potential earnings. We focus on the identification of the non-pecuniary component, which is key to disentangle the relative importance of monetary incentives versus preferences in the context of sorting across sectors. By making the most of the structure of the selection equation, we show that this component is point identified from the knowledge of the covariates effects on earnings, as soon as one covariate is continuous. Notably, and in contrast to most results on the identification of Roy models, this implies that identification can be achieved without any exclusion restriction nor large support condition on the covariates. As a byproduct, bounds are obtained on the distribution of the \textit{ex ante} monetary returns. We also propose a three-stage semiparametric estimation procedure for this model, which yields root-n consistent and asymptotically normal estimators. Finally, 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, \textit{ex ante} returns to schooling.

---

*We thank Victor Aguirregabiria, Christian Belzil, Gerard van den Berg, Federico Bugni, Philippe Février, Marc Gurgand, Marc Henry, Bo Honoré, Shaeekh Khan, Simon Lee, Thierry Magnac, Enno Mammen, Salvador Navarro, David Neumark, Jean-Marc Robin, Christopher Taber, participants at the ESRC Econometric Study Group (July 2009), at the Econometric Society European Meeting (August 2009), at Stats in the Château (September 2009), at the World Congress of the Econometric Society (August 2010) and in seminars at CREST, Bristol, California at Irvine, Duke, Lausanne, Mannheim, Montréal Pompeu Fabra, Rochester, Sciences Po, Toronto and Warwick for helpful discussions and comments at various stages of this research.

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

‡Duke University. E-mail address: apm16@duke.edu.
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, wage returns to education, migration and occupation wage premia. The seminal Roy’s model (1951) of occupational choice can be seen as an extreme setting of self-selection, where agents choose between two sectors by maximizing their 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, however, is 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, Arcidiacono, 2004, and 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 monetary returns and the psychic costs associated with the decision to migrate (see, e.g., Bayer et al., 2010). 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 this _ex ante_ uncertainty, their decision depends on expectations of these potential earnings rather than on their true values.

This paper focuses on the nonparametric identification of the non-pecuniary component in an extended Roy model including these two aspects. Identifying these non-pecuniary factors is key to disentangle the relative importance of monetary incentives versus preferences in the context of sorting across sectors. Our identification strategy proceeds in two stages. We first propose two alternative strategies for identifying the covariates effects on sector-specific earnings. The first one is based on exclusion restrictions. It requires either

---

1The seminal work by Heckman & Honoré (1990) examines the identification of the standard Roy model. See Buera (2006) for an extension to non-separable functional forms for the potential outcomes.
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 builds on an argument at infinity for the potential outcomes, relying on a result from a companion paper (D’Haultfoeuille & Maurel, 2009). This latter approach does not require any exclusion restriction, nor any large support on the covariates. We then consider the identification of the non-pecuniary component of the selection equation. By making the most of the structure imposed on the selection process, we show that this non-pecuniary component is point identified from the knowledge of the covariates effects on earnings, as soon as one covariate is continuous. When all covariates are discrete, our strategy can be naturally adapted to yield informative bounds. The non-pecuniary component can therefore be identified without assuming access to a covariate shifting the selection probability but not the potential outcomes. Combined with our identification at infinity argument, this result implies that identification can be achieved without any exclusion restriction nor large support condition on the covariates. This contrasts with the prevailing view in the literature dealing with the identification of Roy models (see, e.g., French & Taber, 2011).

As a byproduct of this analysis, we obtain informative bounds on the distribution of the ex ante returns, which correspond to the monetary returns expected by the agent at the time of the choice and are also equal, in our setting, to marginal treatment effects evaluated at certain margins (see Heckman & Vytlacil, 2005). We also provide support conditions under which these bounds shrink to a point. In particular, standard average treatment effect parameters are point identified if the probability of selection ranges from zero to one, a result in line with that of Heckman & Vytlacil (2005) in the case of local instrumental variable strategies.

In a recent article investigating the identification of an extended Roy model with a focus on non-pecuniary factors, Bayer et al. (2010) also propose a strategy which does not require any exclusion restriction nor large support condition. However, unlike the model we consider in this paper, Bayer et al. (2010) specify an extended Roy model which does not account for ex ante uncertainty on the outcomes and restrict the alternative-specific non-pecuniary factors to be constant across individuals. Their model also differs from ours in that they consider a setting with potentially more than two sectors, so that our framework does not nest their model. They show that the non-pecuniary factors associated with each choice alternative as well as the unconditional wage distributions are identified provided

---

2 The extended Roy model we consider in the paper allows the non-pecuniary component to vary across individuals according to observed covariates. This differentiates the extended from the generalized Roy model, which allows for unobserved determinants of this component.
that the distribution of monetary returns has a finite lower bound. Although appealing in
that it does not require any exclusion restriction no strong support condition, the finite
lower bound condition may be restrictive and the strategy hard to apply in practice, no-
tably when using log wages which do not have a natural lower bound, as for instance in
Willis & Rosen (1979) and in our application.3

Apart from identification, we also propose a three-stage semiparametric estimation proce-
dure under an index restriction on the effects of covariates. The first two stages allow us
to estimate the covariates effects on potential earnings and correspond to Newey’s method
(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 compo-
nent. This stage simply amounts to estimate a linear instrumental model. The difference
with a standard IV linear model is that both the dependent variable and one of the re-
gressors have to be estimated, this involving in particular a nonparametric 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 reasonably well in finite samples.

Eventually, we apply our estimation procedure to the context of higher education atten-
dance decisions in France over the nineties. We estimate semiparametrically 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 could be expected, we cannot reject at the 10% level the
hypothesis that the local average income for high school graduates only affects the proba-
bority of attendance through the \textit{ex ante} returns to higher education. This allows us to apply
a constrained version of our 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. We find in particular that
10% of the individuals attending higher education choose to do so in spite of negative \textit{ex}
\textit{ante} monetary returns to education. Besides, it follows from our estimates that the higher
education attendance rate would fall from 83.1% to 72% if non-pecuniary factors did not
exist. This decrease is eight time larger than the one associated with a 10% permanent
decrease in labor market earnings of higher education attendees.

The remainder of the paper is organized as follows. Section 2 presents the extended Roy

3Bayer et al. (2010) alternatively prove identification assuming independence between the potential
wages. We do not make this assumption in the paper.
model which is considered throughout the paper and derives our identification results
for the covariates effects on earnings, the non-pecuniary component and the distribution
of the \textit{ex ante} returns. 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 Monte
Carlo simulations. Section 5 applies the preceding estimators to investigate 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 \textit{ex ante} uncer-
tainty 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
$(\eta_0, \eta_1)$ are sector-specific productivity terms known by the agent at the time of the choice
but unobserved by the econometrician. We maintain the following assumption throughout
the paper.

\textbf{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 independence assumption ($X \perp \perp (\eta_0, \eta_1)$) is commonly made when studying sample
selection models (see, e.g., Powell, 1994) or Roy models (see, e.g., Heckman & Honoré,
1990, for the standard Roy model and French & Taber, 2011, for generalized Roy models).
We shall discuss further in the paper how this assumption could be weakened.

We let hereafter $\nu_k = Y_k - E(Y_k | X, \eta_0, \eta_1)$ denote the unexpected shock on $Y_k$ and $\varepsilon_k = \eta_k + \nu_k$ denote the sector-specific residual.\textsuperscript{4} Note-
worthy, 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

\textsuperscript{4}Part of the residual $\nu_k$ 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.
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 \textit{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 \(U_k = E(Y_k|X, \eta_0, \eta_1) + G_k(X)\). \(U_k\) 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 \(X\). Assuming additive separability between the expected earnings and the non-pecuniary component of utility is standard for the generalizations of the Roy model considered in the literature.\(^5\) This separability assumption, which is required to obtain an additive separable form between \(X\) and \(\eta_{\Delta}\) in the selection index, is key for our identification strategy.\(^6\) Along with the covariates \(X\), the econometrician observes the chosen sector \(D\), which satisfies

\[
D = \mathbb{1}\{U_1 > U_0\} = \mathbb{1}\{\eta_{\Delta} > \psi_0(X) - \psi_1(X) + G(X)\},
\]

(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 known in the literature as the extended Roy model, whose identification is also considered, in a version without \textit{ex ante} uncertainty, by Heckman & Vytlacil (2007). Bayer et al. (2010) examine the identification of an extended Roy model without \textit{ex ante} uncertainty as well, which allows for more than two sectors and includes a non-pecuniary intercept for each sector. In a recent paper, Fox & Gandhi (2011) extend this model by allowing for random functions in the selection equation.\(^7\) The model presented above 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).

A central contribution of this paper is to show that, by making the most of the extended

\(^5\)Note that \(-G_k(X)\) can be seen as a cost of entry into sector \(k\). This interpretation is put forward in the treatment effect literature relying on generalized Roy models.

\(^6\)This echoes the fact that additive separability in the selection index is crucial for the identification results obtained in the Marginal Treatment Effects literature (see, e.g., Heckman & Vytlacil, 2005).

\(^7\)However, as is the case for the model we consider, Fox & Gandhi (2011) rule out the existence of additive errors for the non-pecuniary components entering the selection model.
Roy structure, the identification of the covariates effects on earnings directly entails the identification of the non-pecuniary component. In particular, unlike in Heckman & Vytlacil (2007) and Fox & Gandhi (2011), no exclusion restriction between $G$ and $(\psi_0, \psi_1)$ is needed.

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 $\eta_\Delta$ admits a density, denoted by $f_{\eta_\Delta}$, with respect to the Lebesgue measure.

Assumption 2.2 is an innocuous normalization which stems from the fact that adding a constant to $\psi_k$ and subtracting it to $\eta_k$ does not modify the model. Assumption 2.3 is a technical condition which is usual in Roy or competing risks 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 $(\psi_0, \psi_1)$. It is important to note that the following results only rely on the reduced form version of the model. The structure imposed on the selection equation will be crucial to identify the non-pecuniary component and the distribution of the *ex ante* returns, but is not needed at this stage. The first and standard approach we focus on is based on exclusion restrictions, in the same spirit as, e.g., Das et al. (2003). The second hinges on a nonstandard identification at infinity, with the advantage of not requiring any exclusion restriction. The first strategy relies on the following assumption.

**Assumption 2.4 (Exclusion restrictions)** $\psi_0$ (resp. $\psi_1$) depends only on $\tilde{X}_0 \subset X$ (resp. on $\tilde{X}_1 \subset X$). Moreover, $\tilde{X}_0$ (resp. $\tilde{X}_1$) and $P(D = 1|X)$ are measurably separated, that is, any function of $\tilde{X}_0$ (resp. of $\tilde{X}_1$) 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 $X = (\tilde{X}_0, Z)$ and $\tilde{X}_1 = \tilde{X}_0$. 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 use sector-specific regressors in our application.

Intuitively, the measurable separation requirement\(^8\) 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,
also made by Das et al. (2003), 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 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 \( \psi_0 \) and \( \psi_1 \) up to location parameters. Then full identification stems
from the normalization of Assumption 2.2. Similarly to Das et al. (2003), Proposition 2.1
below does not provide any result on the location parameters. In general, such parameters
are identified only at infinity under a large support condition, i.e. when \( P(D = 1|X) \)
can be arbitrarily close to zero and one (see Heckman, 1990). Finally, following Das et al.
(2003), one could actually relax the independence condition \( X \perp \eta \) and allow for
endogenous covariates \( X \) while still identifying \( \psi_0 \) and \( \psi_1 \) up to location. However, it is
not clear in this case how to recover the non-pecuniary component \( G(X) \).

**Proposition 2.1** Suppose that Assumptions 2.1-2.4 hold. Then \( \psi_0 \) and \( \psi_1 \) are identified.

The main difficulty in the proof of Proposition 2.1 is to establish that \( E(\varepsilon_k|D = k, X) \)
only depends on \( P(D = 1|X) \), especially because the distribution of \( \eta_\Delta \) may not be strictly
increasing everywhere. Once this is established, we can rely on the measurable separability
condition of Assumption 2.4 to prove the result.\(^9\)

\(^8\)We adopt here the terminology of Florens et al. (2008) (see their Assumption A4).

\(^9\)Unlikely in Das et al. (2003), identification is shown without assuming that the regressors are continuous
nor that \( \psi_0 \) and \( \psi_1 \) are continuously differentiable.
Alternatively, $\psi_0$ and $\psi_1$ can also be identified at the limit without any exclusion restriction, under the following restrictions on the error terms.

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

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

The first restriction reinforces the condition that $X \perp (\eta_0, \eta_1)$, by ruling out in particular heteroskedasticity of the shocks $(\nu_0, \nu_1)$. The second restriction is a light tail condition, which is in practice fairly mild. The last one can be interpreted as a moderate dependence condition between $\eta_0$ and $\eta_1$, which is not very restrictive either. When $(\eta_0, \eta_1, \nu_0, \nu_1)$ is gaussian for instance, one can show that it is equivalent to $\text{cov}(\eta_0, \eta_1) < \min(V(\eta_0), V(\eta_1))$. In particular, when $V(\eta_0) = V(\eta_1)$, this condition is automatically satisfied, except in the degenerate case where $\eta_0 = \eta_1$.

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

Proposition 2.2 is based on a result by D’Haultfoeuille & Maurel (2009), and on the fact that under Assumption 2.5,

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

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 used 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 extended Roy models.

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 $X$. 

---

10 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.  
In particular, it may be applied even if $X$ is discrete. 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 $(\psi_0, \psi_1)$, and that Assumption 2.3 holds. Contrary to previously, we fully rely here on the detailed structure of the model, and in particular on the link between the residuals in the outcome equations and the one in the selection equation. We first suppose that at least one of the components $X_j$ of $X$, say $X_1$, is continuous, and impose a mild regularity condition on $(T + G)(.)$, where $T(.) = \psi_0(.) - \psi_1(.)$.

#### Assumption 2.6

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

We start from the following observations:

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

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

Let $X_{-1} = (X_2, ...)$, so that $X = (X_1, X_{-1})$ (and we let similarly $x = (x_1, x_{-1})$). Under Assumptions 2.3 and 2.6, the functions $g_0(x) = E(D|X = x)$ and $E[D\eta_\Delta|X = x]$ are differentiable with respect to $x_1$. Thus, Equations (2.3) and (2.4) imply that

$$\frac{\partial E[D\eta_\Delta|X = x]}{\partial x_1} = (T(x) + G(x))\frac{\partial q_0}{\partial x_1}(x).$$

Because $T(.)$ and $g_0(.)$ are identified, this equation shows that, provided that $\partial q_0/\partial x_1(x) \neq 0$, identification of $G(x)$ amounts to recovering $\partial E[D\eta_\Delta|X = x]/\partial x_1$. The key idea, for that purpose, is to relate this term with the residual $\varepsilon$ of the (realized) outcome equation. Observe that by definition of $\nu_i$ and the law of iterated expectations, $E(\nu_k|D = k, X) = 0$. As a result, letting $\varepsilon = D\varepsilon_1 + (1 - D)\varepsilon_0$, we get

$$E(\varepsilon|X) = E[D\varepsilon_1 + (1 - D)\varepsilon_0|X]$$

$$= E[D\eta_1 + (1 - D)\eta_0|X]$$

$$= E[D\eta_\Delta|X] + E[\eta_0]. \quad (2.5)$$

\[12\] If one of the covariates has large support, one can use alternatively the results of Lewbel (2007) which also yield identification of the covariates effects on earnings without any instrument for selection.
Thus, letting $g_0(x) = E(\varepsilon | X = x)$, we obtain
\[ \frac{\partial g_0}{\partial x_1}(x) = (T(x) + G(x)) \frac{\partial q_0}{\partial x_1}(x). \] (2.6)

Since $\varepsilon = Y - \psi_D(X)$ is identified, $g_0$ and $q_0$ are identified and we can use Equation (2.6) to recover $G(.)$. The only exception is actually 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 point identified.\(^{13}\)

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

The independence condition between $X$ and $(\eta_0, \eta_1)$ plays an important role in the derivation above. However, this assumption could be weakened to the conditional independence condition $X_1 \perp \perp (\eta_0, \eta_1)|X_-1$, without affecting the identification result. We maintain the stronger independence assumption here for the sake of notational simplicity.

Now consider the case where no component of $X$ is continuous, so that $X$ has a discrete distribution. Suppose that it takes $M < \infty$ 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 $G$, replacing derivatives with finite differences. First, note that $P(D = 0 | X = x) = F_{\eta}(T(x) + G(x))$, with $F_{\eta}$ denoting the cumulative distribution function of $\eta$. This equality implies that we can sort the $x_i$’s so that $T(x_i) + G(x_i) < ... < T(x_M) + G(x_M)$.\(^{14}\) This provides a first set of inequalities on $(G(x_1), ..., G(x_M))$. Besides, letting $i < j$, we have,

\[
\sum_{k=i}^{j-1} [T(x_{k+1}) + G(x_{k+1})] [g_0(x_{k+1}) - g_0(x_k)] \\
\leq g_0(x_j) - g_0(x_i) = - \int_{T(x_i) + G(x_i)}^{T(x_j) + G(x_j)} u f_{\eta}(u) du \\
\leq \sum_{k=i}^{j-1} [T(x_k) + G(x_k)] [g_0(x_{k+1}) - g_0(x_k)].
\]

\(^{13}\)If the second condition of Assumption 2.7 fails to hold, $\frac{\partial G}{\partial x_1}$ is still identified (but not $G$), as it is equal to $-\frac{\partial G}{\partial x_1}$ in this case. Besides, since 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.

\(^{14}\)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 $G(x_i)$. 

11
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 \(G(x_2), ..., G(x_{M-1})\).

In the following, consistent with the framework of our application, we will maintain the assumptions ensuring that the non-pecuniary component \(G(.)\) is point identified. We leave in particular the analysis of set-estimation of \(G(.)\) for further research.

2.4 Distribution of \textit{ex ante} returns

We now turn to the identification of the distribution of the \textit{ex ante} returns, \(\Delta = E(Y_1 - Y_0|X, \eta_0, \eta_1)\). The \textit{ex ante} return is meaningful since it corresponds to what agents act on (see Cunha & Heckman, 2007). Besides, it corresponds to the \textit{ex post} return 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. Although we have remained completely agnostic on the information set of the agents, it is possible to point or partially identify the distribution of \(\Delta\). The intuition behind is similar to that underlying the identification of \(G(.)\). \(\Delta\) depends on \(\eta_\Delta\), which is also the residual of the selection equation. Thus, the observed choice of sector directly provides information on these \textit{ex ante} returns. To see this, first recall that

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

This shows that \(F_{\eta_\Delta}\) is identified over the support of \(T(X) + G(X)\). Now, the cumulative distribution function of \(\Delta\) (\(F_\Delta\)) satisfies

\[
F_\Delta(u) = E[P(\eta_\Delta \leq u + T(X)|X)] = E[F_{\eta_\Delta}(u + T(X))].
\]

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 \textit{ex ante} returns \(\Delta\) 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 (i.e. for the individuals such that \(D = 1\) here), by integrating the \textit{ex ante} returns over the distribution of \(\eta_\Delta\). But even if this large support condition fails, it is still possible to point identify a subset of the distribution of the \textit{ex ante} returns, and bound \(F_\Delta(u)\) for the rest of the distribution.\(^{15}\) Indeed, letting \([M, \overline{M}]\) (resp. \([P, \overline{P}]\))

\(^{15}\)Heckman & Vytlacil (2007) also obtain bounds on the average returns without assuming large support on the selection probability, in the context of an extended Roy model. Their strategy hinges on an exclusion
denote the support of $T(X) + G(X)$ (resp. of $P(D = 0|X)$), we have, by the monotonicity of $F_{\eta \Delta}$, $F_{\Delta}(u) \in [F_{\Delta}(u), \overline{F}_{\Delta}(u)]$, where

\[
F_{\Delta}(u) = E \left( F_{\eta \Delta} (u + T(X)) \mathbb{1} \{u + T(X) \in [M, \overline{M}]\} \right) + \overline{P} \times P(u + T(X) > \overline{M}) + 0 \times P(u + T(X) \leq M), \tag{2.7}
\]

\[
\overline{F}_{\Delta}(u) = E \left( F_{\eta \Delta} (u + T(X)) \mathbb{1} \{u + T(X) \in [M, \overline{M}]\} \right) + 1 \times P(u + T(X) > \overline{M}) + P \times P(u + T(X) \leq M). \tag{2.8}
\]

The distribution of the \textit{ex ante} treatment effect on the treated can be identified in a similar way, with

\[
F_{\Delta|D=1}(u) = \frac{E \{ (F_{\eta \Delta} (u + T(X)) - P(D = 0|X)) \times \mathbb{1} \{G(X) \leq u\} \} }{P(D = 1)}. \tag{2.9}
\]

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

\[
\Delta^{MTE}(x,p) = E(Y_1 - Y_0 | X = x, S_{\eta \Delta}(\eta \Delta) = p) = \psi_1(x) - \psi_0(x) + S^{-1}_{\eta \Delta}(p)
\]

Thus, $\Delta = (\psi_1 - \psi_0)(X) + \eta \Delta$ coincides with $\Delta^{MTE}(X, S_{\eta \Delta}(\eta \Delta))$. Besides, one is able to identify $\Delta^{MTE}(x,p)$ for all $p$ in the support of $P(D = 1|X)$, since in that case there exists $\tilde{x}$ in the support of $X$ such that $S^{-1}_{\eta \Delta}(p) = (\psi_0 - \psi_1 + G)(\tilde{x})$.

### 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 $\psi_0(\cdot), \psi_1(\cdot)$ and $G(\cdot)$. More precisely, we consider extended Roy models with a linear index structure of the form:\textsuperscript{16}

\[
\begin{align*}
Y_0 &= X'\beta_0 + \varepsilon_0 \\
Y_1 &= X'\beta_1 + \varepsilon_1 \\
D &= \mathbb{1} \{-\delta_0 + X'(\beta_1 - \beta_0 - \gamma_0) + \eta \Delta > 0\}.
\end{align*}
\tag{3.1}
\]

\textsuperscript{16}We suppose without loss of generality that the constant is not included in $X$, so that $\varepsilon_0$ and $\varepsilon_1$ do not necessarily have mean zero.

13
In this setting, the non-pecuniary component $G(X)$ is of the form $\delta_0 + X'\gamma_0$. Let $\gamma_{0j}$ (resp. $\beta_{0j}$, $\beta_{1j}$) denote the $j$-th component of $\gamma_0$ (resp. $\beta_0$, $\beta_1$). We also impose the following conditions.

**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 require 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 $\tilde{\eta}_{\Delta} = (\eta_{\Delta} - \delta_0)/(\beta_{11} - \beta_{01} - \gamma_{01})$. We propose a three-stage estimation procedure of the model, where we estimate first $\zeta_0$, then $(\beta_0, \beta_1)$ and finally $(\delta_0, \gamma_0)$. In line with the identification section, the first and second stages of our procedure are not new, and rely on the fact that we can rewrite the model in the following reduced form:

$$
D = I\{X'\zeta_0 + \tilde{\eta}_{\Delta} > 0\} \quad (3.2)
$$

$$
Y_k = X'\beta_k + \varepsilon_k, \quad k \in \{0, 1\},
$$

where $Y_k$ is observed when $D = k$, $\tilde{\eta}_{\Delta}$ is independent of $X$ and $E(\varepsilon_k|D = k, X)$ only depends on $X'\zeta_0$.\footnote{Indeed, $\varepsilon_k = \eta_k + \nu_k$ with $E(\nu_k|D = k, X) = 0$ by definition and $E(\eta_1|D = 1, X = x) = E(\eta_1|\tilde{\eta}_{\Delta} > -x'\zeta_0)$ (and similarly for $k = 0$). Note that in general, $\varepsilon_k$ is not independent of $X$ because $\nu_k$ is not.}

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 \( \zeta_0 \) by a single index estimator \( \hat{\zeta} \), for which we suppose Assumption 3.4 to be satisfied. This is the case of many semiparametric estimators, such as the one of Klein & Spady (1993) or Ichimura (1993). Secondly, we estimate \( \beta_0 \) and \( \beta_1 \) by series estimator, and we suppose that they satisfy 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

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

**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

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

The originality of the estimation procedure lies in its third stage, which is devoted to the estimation of \((\delta_0, \gamma_0)\). Actually, it suffices to estimate \( \delta_0 \) 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:

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

where \( \tilde{\lambda}_0 \) 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 \( \lambda_0 = \tilde{\lambda}_0 - u_0q_0(u_0)\alpha_0 \). 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 \( \varepsilon = W'\theta_0 + \xi \). The regressors \( D \) and \( V \) are endogenous since selection \( D \) depends both on \( U \) and \( \eta_\Delta \). We therefore use an IV estimator of \( \theta_0 \) with functions of the index \( U \) as instruments for \( D \) and \( V \). To avoid boundary effects, we include some trimming by considering feasible versions of the instruments \( Z = \mathbb{1}\{X \in \mathcal{X}\}h(U) \), 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 \( U \).\(^{18}\) Then \( \theta_0 = E(ZW')^{-1}E(Z\varepsilon) \), and we estimate it by

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

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

\[
\hat{V}_i = D_i\hat{U}_i - \int_{u_0}^{D_i} \hat{q}(u, \hat{\zeta})du,
\]

\[
\hat{Z}_i = \mathbb{1}\{X_i \in \mathcal{X}\}h\left( \hat{U}_i \right).
\]

Finally, \( \hat{U}_i = X_i'\hat{\zeta} \) and

\[
\hat{q}(u, \zeta) = \frac{\sum_{i=1}^{n} D_iK\left( \frac{u - X_i'\zeta}{h_n} \right)}{\sum_{i=1}^{n} K\left( \frac{u - X_i'\zeta}{h_n} \right)}.
\]

(3.6)

where \( K(.) \) is a kernel function and \( h_n \) the bandwidth parameter. The result on the third step estimator \( \hat{\theta} \) 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 \( \hat{Z}_i - Z_i \) 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

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

\(^{18}\)This trimming procedure ensures uniform consistency of kernel estimators over \( \{x'\zeta_0, x \in \mathcal{X}\} \).
where $\Omega_{11}$ is defined by Equation (7.8) in Appendix A and

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

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

Theorem 3.1 establishes the root-n consistency and asymptotic normality of $\hat{\theta}$. We prove the result by first remarking that $\hat{\theta}$ is a two-step GMM estimator with a nonparametric first step estimator ($\hat{q}$). We then follow Newey & McFadden (1994)’s outline for establishing asymptotic normality. Some differences arise however because $\hat{q}$ also depends on the estimator $\hat{\zeta}$. Theorem 3.1 also shows that the asymptotic variance of $\hat{\theta}$ depends on the three variables $\Omega_{11}$, $\Omega_{21}$ 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 $\hat{V}_i$. The third one corresponds to the moment estimation of the linear instrumental model (3.5) in the last step.

As the proof of the theorem shows, $\hat{\theta}$ can be linearized. Thus, by Assumptions 3.4 and 3.5, the estimator of $\gamma_0$, $\hat{\gamma} = \hat{\beta}_1 - \hat{\beta}_0 + \hat{\alpha} \hat{\zeta}$, is also root-n consistent and asymptotically normal.

Although $\delta_0$ and $\gamma_0$ are identified without any exclusion restriction, imposing restrictions on $\gamma_0$ 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 $\gamma_{01} = 0$. In this case, we get from the second stage $\alpha_0 = \beta_{01} - \beta_{11}$. Hence, $\gamma_0 = \beta_1 - \beta_0 + \alpha_0 \zeta_0$ 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 $\delta_0$ only, through the instrumental linear model

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

where $\alpha_0$ in the left hand side can now be estimated by $\hat{\beta}_{01} - \hat{\beta}_{11}$. One can show that the corresponding estimator is also asymptotically normal.\footnote{The proof is very close to the one of Theorem 3.1 and is available from the authors upon request.}

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, for four...
different sample sizes (namely \( n = 500, n = 1,000, n = 2,000 \) and \( n = 10,000 \)):

\[
Y_{0i} = X_{2i}\beta_{02} + X_{3i}\beta_{03} + \eta_{0i} + \nu_{0i}
\]

\[
Y_{1i} = X_{1i}\beta_{11} + X_{3i}\beta_{13} + \eta_{1i} + \nu_{1i}
\]

\[
D_i = 1 \{ -\delta_0 + X_{1i}(\beta_{11} - \gamma_{01}) + X_{2i}(\beta_{02} - \gamma_{02}) + X_{3i}(\beta_{13} - \beta_{03} - \gamma_{03}) + \eta_{1i} - \eta_{0i} > 0 \}.
\]

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 \( p = 0.5 \). We let \( (\eta_{0i}, \eta_{1i})' \) be joint normal, with mean \((0, 1)\)' and a variance \( \Sigma \) such that \( \Sigma_{11} = \Sigma_{22} = 1 \) and \( \Sigma_{12} = \Sigma_{21} = 0.5 \). \((\nu_{0i}, \nu_{1i})'\) are drawn from a heteroskedastic normal distribution, with 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 \( \beta_0 \) and \( \beta_1 \). The series estimator of the selection correction term was computed using the inverse Mills ratio transform and Legendre polynomials of order increasing with the sample size at a rate \( n^{1/8} \) (see Newey, 2009, Equation (3.6) and Assumption (4.5) respectively).

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 \( \delta_0 \) and \( \gamma_0 \) with the quartic kernel suggested in Section 3 and a bandwidth \( h_n = 0.5\sigma(\hat{U})n^{-1/7} \), where \( \sigma(\hat{U}) \) is the estimated standard deviation of \( \hat{U} \). We choose the functions \( h_1(x) = \Phi(\hat{a}_0 + \hat{a}_1 x) \) and \( h_2(x) = x h_1(x) - \int_{\hat{a}_0}^{\hat{U}} q(u, \hat{\zeta}) du \) for the instruments, where \( \Phi(.) \) denotes the normal cumulative distribution, \((\hat{a}_0, \hat{a}_1)\) is the probit estimator of \( D \) on \((1, \hat{U})\) and \( \hat{u}_0 \) is the sample minimum of \( \hat{U} \).

Finally, no trimming was performed since it did not improve the accuracy of the estimators in our setting.

The performance of the estimators are summarized in Panel A of Table 1 below, which reports for each parameter its bias, standard deviation and root mean squared error (RMSE).

\(^{20}\)Namely, polynomials of order 6 are used for sample sizes \( n = 500 \) and \( n = 1,000 \), order 7 for \( n = 2,000 \) and 8 for \( n = 10,000 \).

\(^{21}\)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.
The non-pecuniary components $\gamma_0$ and $\delta_0$ are less precisely estimated and the corresponding estimators display a larger bias than that of the outcome equations parameters, $\beta_0$ and $\beta_1$. Basically, this stems from the sequential structure of the proposed estimation procedure, and from the fact that, in this specification, no exclusion restriction is used to identify $\gamma_0$ and $\delta_0$. Despite this, the estimators of $\gamma_0$ and $\delta_0$ are accurate enough for being able to reject in most simulation samples the hypothesis that $\delta_0 = 0$ or $\gamma_{03} = 0$ for $n = 2,000$ and $n = 10,000$, as well as the hypothesis that $\gamma_{02} = 0$ for $n = 10,000$. Besides, even for small sample sizes, these estimators display a negligible bias, compared to their standard deviation, which is reassuring for conducting valid inference.

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 $\gamma_0$ in the former case with the first step estimates alone, and use Equation (3.7) to estimate $\delta_0$ only. The properties of the unconstrained and constrained estimators are displayed respectively in Panel B and C of Table 1. Using an exclusion restriction on the non-pecuniary component leads to a substantial improvement in the performances of the estimators of $\gamma_0$, for all sample sizes. In particular, the standard error for $\gamma_{02}$ decreases by about 80% between the two specifications. The performance of the estimator for $\delta_0$, which is still estimated in a third step, is very similar to the unconstrained specification. Overall, it appears from these Monte Carlo simulations that the constrained estimator should be preferred in the presence of an exclusion restriction on the non-pecuniary component. Importantly, whether such an exclusion restriction is valid can be directly tested in the data after estimating the unconstrained model.
<table>
<thead>
<tr>
<th>$n$</th>
<th>Coeff.</th>
<th>Panel A</th>
<th>Panel B</th>
<th>Panel C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bias</td>
<td>Std dev</td>
<td>RMSE</td>
</tr>
<tr>
<td>500</td>
<td>$\beta_{02}$</td>
<td>-0.009</td>
<td>0.174</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>$\beta_{03}$</td>
<td>-0.011</td>
<td>0.288</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>0.003</td>
<td>0.136</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>$\beta_{13}$</td>
<td>-0.001</td>
<td>0.172</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{01}$</td>
<td>0.083</td>
<td>1.781</td>
<td>1.783</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{02}$</td>
<td>-0.106</td>
<td>1.033</td>
<td>1.038</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{03}$</td>
<td>0.035</td>
<td>0.475</td>
<td>0.476</td>
</tr>
<tr>
<td></td>
<td>$\delta_0$</td>
<td>0.090</td>
<td>0.747</td>
<td>0.753</td>
</tr>
</tbody>
</table>

|     | $\beta_{02}$ | -0.009 | 0.124  | 0.124  | -0.006 | 0.110  | 0.110  | -0.006 | 0.110  | 0.110  |
|     | $\beta_{03}$ | -0.003 | 0.206  | 0.206  | 0.002  | 0.180  | 0.180  | 0.002  | 0.180  | 0.180  |
|     | $\beta_{11}$ | 0.005  | 0.089  | 0.089  | -0.001 | 0.086  | 0.086  | -0.001 | 0.086  | 0.086  |
|     | $\beta_{13}$ | 0.001  | 0.122  | 0.122  | 0.008  | 0.127  | 0.127  | 0.008  | 0.127  | 0.127  |
|     | $\gamma_{01}$ | -0.021 | 1.201  | 1.201  | 0.063  | 0.879  | 0.881  | (not estimated) |
|     | $\gamma_{02}$ | -0.012 | 0.695  | 0.695  | -0.067 | 0.651  | 0.655  | -0.019 | 0.140  | 0.141  |
|     | $\gamma_{03}$ | 0.020  | 0.323  | 0.323  | 0.013  | 0.297  | 0.297  | 0.006  | 0.251  | 0.251  |
|     | $\delta_0$ | 0.022  | 0.510  | 0.510  | 0.037  | 0.454  | 0.456  | 0.031  | 0.446  | 0.447  |

|     | $\beta_{02}$ | -0.005 | 0.090  | 0.090  | 0.000  | 0.080  | 0.080  | 0.000  | 0.080  | 0.080  |
|     | $\beta_{03}$ | -0.001 | 0.136  | 0.136  | 0.003  | 0.125  | 0.125  | 0.003  | 0.125  | 0.125  |
|     | $\beta_{11}$ | 0.001  | 0.067  | 0.067  | 0.002  | 0.062  | 0.062  | 0.002  | 0.062  | 0.062  |
|     | $\beta_{13}$ | -0.001 | 0.091  | 0.091  | 0.002  | 0.091  | 0.091  | 0.002  | 0.091  | 0.091  |
|     | $\gamma_{01}$ | 0.047  | 0.883  | 0.884  | 0.031  | 0.628  | 0.629  | (not estimated) |
|     | $\gamma_{02}$ | -0.039 | 0.517  | 0.519  | -0.036 | 0.466  | 0.467  | -0.014 | 0.100  | 0.101  |
|     | $\gamma_{03}$ | 0.021  | 0.233  | 0.234  | 0.006  | 0.210  | 0.210  | 0.004  | 0.184  | 0.184  |
|     | $\delta_0$ | 0.012  | 0.360  | 0.360  | 0.023  | 0.332  | 0.333  | 0.023  | 0.331  | 0.332  |

|     | $\beta_{02}$ | 0.003  | 0.039  | 0.039  | -0.002 | 0.034  | 0.034  | -0.003 | 0.036  | 0.036  |
|     | $\beta_{03}$ | -0.005 | 0.067  | 0.067  | 0.001  | 0.057  | 0.057  | 0.002  | 0.057  | 0.058  |
|     | $\beta_{11}$ | -0.003 | 0.028  | 0.028  | 0.001  | 0.026  | 0.027  | 0.000  | 0.027  | 0.027  |
|     | $\beta_{13}$ | -0.002 | 0.039  | 0.039  | -0.004 | 0.041  | 0.041  | -0.001 | 0.041  | 0.041  |
|     | $\gamma_{01}$ | 0.084  | 0.385  | 0.394  | 0.015  | 0.276  | 0.276  | (not estimated) |
|     | $\gamma_{02}$ | -0.024 | 0.232  | 0.233  | -0.010 | 0.204  | 0.204  | 0.000  | 0.043  | 0.043  |
|     | $\gamma_{03}$ | 0.012  | 0.104  | 0.105  | -0.001 | 0.088  | 0.088  | -0.005 | 0.079  | 0.079  |
|     | $\delta_0$ | -0.049 | 0.153  | 0.161  | -0.004 | 0.136  | 0.137  | 0.000  | 0.141  | 0.142  |

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 $(\delta_0, \gamma_0)$ 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
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. After completing secondary education, individuals are assumed to decide either to enter directly the labor market with a high school degree ($k = 0$) or to attend higher education ($k = 1$). They are supposed to make their decision $D \in \{0, 1\}$ by comparing the expected discounted 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 $X$, $(\eta_0, \eta_1)$ are individual productivity terms which are supposed to be known by the individual at the time of her decision but unobserved by the econometrician and $(\nu_0, \nu_1)$ represent random shocks with means zero, which are unobserved by both the individual and the econometrician.

---

22 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 different framework based on a factor structure on the unobservables. 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).

23 The French higher education system includes universities, which do not impose any entry selection, as well as the Grandes Ecoles and specialized technical colleges, which are selective.
The expected utility $U_k$ of each schooling decision $k$ is supposed to be given by

$$U_k = E(Y^*_k|X, \eta_k) + G_k(X),$$

where $G_k(X)$ denotes the consumption value associated with the schooling decision $k$. 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 $G_0 - G_1$, 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 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. 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. These two databases 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 datasets, we exploit the pooled dataset hereafter.

---

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

25 Beffy et al. (2010) also use these data to estimate the influence of expected returns when choosing a college major.
### Table 2: Descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Higher education attendees</th>
<th>High school level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
</tr>
<tr>
<td>Initial monthly log wage (1992 French Francs)</td>
<td>8.75</td>
<td>0.44</td>
</tr>
<tr>
<td>Secondary schooling track</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L (Humanities)</td>
<td>0.15</td>
<td>0.36</td>
</tr>
<tr>
<td>ES (Economics and Social Sciences)</td>
<td>0.17</td>
<td>0.38</td>
</tr>
<tr>
<td>S (Sciences)</td>
<td>0.32</td>
<td>0.47</td>
</tr>
<tr>
<td>Vocational</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>Technical</td>
<td>0.32</td>
<td>0.46</td>
</tr>
<tr>
<td>Born abroad</td>
<td>0.02</td>
<td>0.16</td>
</tr>
<tr>
<td>Father born abroad</td>
<td>0.11</td>
<td>0.32</td>
</tr>
<tr>
<td>Mother born abroad</td>
<td>0.10</td>
<td>0.31</td>
</tr>
<tr>
<td>Entering the labor market in 1992</td>
<td>0.46</td>
<td>0.50</td>
</tr>
<tr>
<td>Entering the labor market in 1998</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td>Male</td>
<td>0.47</td>
<td>0.5</td>
</tr>
<tr>
<td>Father’s profession</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>0.06</td>
<td>0.25</td>
</tr>
<tr>
<td>Tradesman</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>Executive</td>
<td>0.26</td>
<td>0.44</td>
</tr>
<tr>
<td>Intermediate occupation</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>Blue collar</td>
<td>0.17</td>
<td>0.38</td>
</tr>
<tr>
<td>White collar</td>
<td>0.21</td>
<td>0.41</td>
</tr>
<tr>
<td>Other</td>
<td>0.06</td>
<td>0.24</td>
</tr>
<tr>
<td>Age in 6th grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 10</td>
<td>0.10</td>
<td>0.29</td>
</tr>
<tr>
<td>11</td>
<td>0.84</td>
<td>0.37</td>
</tr>
<tr>
<td>≥ 12</td>
<td>0.07</td>
<td>0.25</td>
</tr>
<tr>
<td>Paris region</td>
<td>0.16</td>
<td>0.36</td>
</tr>
<tr>
<td>Number of higher education years</td>
<td>2.82</td>
<td>1.45</td>
</tr>
<tr>
<td>Dropout rate</td>
<td>0.16</td>
<td>0.37</td>
</tr>
<tr>
<td>Number of observations</td>
<td>19,143</td>
<td>5,082</td>
</tr>
</tbody>
</table>

Our subsample of interest comprises respondents having at least passed the national high school final examination. The labor market participation rate for this subsample is above 90% over the period of interest, for both genders. 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 we do not observe any wage for them. This finally leaves us with a large sample of 24,225 individuals. Working with
many observations is especially important for the semiparametric estimation procedure to perform well.\textsuperscript{26} 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 $\psi_0(\cdot), \psi_1(\cdot)$ and $G(\cdot)$ are assumed to depend on the secondary schooling (i.e. high school) track, whether the student is born abroad (and similarly for her parents), her year of entry into the labor market (1992 or 1998), gender, parental profession, age in 6\textsuperscript{th} grade (i.e., her age of entry into junior high school)\textsuperscript{27} and a dummy for living in Paris region (at the time of entry into junior high school). Given that being enrolled in a vocational secondary schooling track seems to be a 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 $\psi_0(\cdot)$ (resp. $\psi_1(\cdot)$) 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 high school and higher education graduates.\textsuperscript{28} 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.

\textsuperscript{26}Papers in this literature usually rely on the NLSY 79 (see Cunha & Heckman, 2007), resulting in samples of around 1,000 observations. Most of these use a flexible parametric estimation procedure.

\textsuperscript{27}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.

\textsuperscript{28}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 \( t \), \( \tau \) denotes the annual discount factor and \( A \) is the duration of active life. We account for the opportunity costs incurred when entering higher education by allowing the year of entry into the labor market \((t_{0,k})\) to vary according to the schooling choice. For a given 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 \( Y^* = D Y_{1}^* + (1 - D) Y_0^* \). Still, we can recover an expectation of this stream of income under additional assumptions on income dynamics. We suppose here that

\[ y_{k,t} = \rho_k \mathbbm{1}\{t - t_{0,k} + 1 \leq B\} + y_{k,t-1} + \nu_{k,t}, \quad (5.1) \]

where \( \rho_k \) 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 independent and identically distributed over time, with mean zero. We introduce the dummy \( \mathbbm{1}\{t - t_{0,k} + 1 \leq B\} \) 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 \( \rho_k \) 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 \( \tilde{\tau}_k = \tau^{t_{0,k}} \left( \frac{1 - \tau^{B+1}}{1 - \tau} \right) \), \( C_k = \tau^{t_{0,k}} \left( \frac{\tau}{(1 - \tau)^2} \right) (1 - \tau^B + B \tau^{A+1} (\tau - 1)) \) and

\[ Y_k = \tilde{\tau}_k y_{D,t_{0,D}} + \rho_k C_k. \]

Because \( \tilde{\tau}_D, C_D \) and \( \rho_D \) are identified for given \( \tau, A \) and \( B \), we can identify \( Y = D Y_1^* + (1 - D) Y_0^* \). 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 \( \rho_0 \) and \( \rho_1 \) 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 Coppejans, 2001) with $K_1 = 3$ mixture components. The second step is performed with Newey (2009)'s series estimator, with $K_2 = 9$ approximating terms, consistently with the growth rate for the order of the polynomials used in the Monte Carlo simulations. We finally use for the last step the same specifications as in the Monte Carlo simulations.

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

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

Therefore, we can obtain an estimator $\hat{F}_{\eta \Delta}(.)$ on $[\hat{M}, \hat{M}]$, the estimated support of $\delta_0 + X'\alpha_0\zeta_0$, by regressing nonparametrically $1 - D$ on the index $\hat{\delta} + X'\hat{\alpha}\hat{\zeta}$. On $[\hat{M}, +\infty)$ (resp. $(-\infty, \hat{M}]$), we simply set estimate $F_{\eta \Delta}(.)$ by $[\hat{P}, 1]$ (resp. $[0, \hat{P}]$), where $\hat{P}$ (resp. $\hat{P}$) is the supremum (resp. infimum) of $\hat{F}_{\eta \Delta}(.)$ on $[\hat{M}, \hat{M}]$. Finally, we estimate $F_\Delta(u)$ and $\bar{F}_\Delta(u)$ with the empirical analogs of (2.7) and (2.8). Bounds on the distribution of the ex ante returns for those attending higher education are estimated similarly, using (2.9). In practice, we consider a kernel estimator of $F_{\eta \Delta}$ with a gaussian kernel, and a bandwidth $\hat{h}_n = 1.6\sigma(\hat{U})n^{-1/5}$.

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 $\beta_0$ and $\beta_1$ 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

---

29 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. Also, the mixture estimator we use here, unlike that of Klein & Spady (1993), is based on an objective function which is bounded, this easing the optimization process.

30 We estimated the model with different values around the baseline tuning parameters $K1, K2$ and the bandwidth $h_n$ used in the estimation of $q_0$ in the third step. The results remain overall stable after such changes.
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 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. 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 clear reason to enter the non-pecuniary factors and should therefore only affect the probability of attendance through the *ex ante* returns. It also indicates that the data is consistent with a constrained specification where the coefficient related to the local average income of high school graduates is set equal to zero.\textsuperscript{31} As already shown in Section 4, our estimation procedure performs 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.\textsuperscript{32} 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

\textsuperscript{31}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 smaller and (ii) the latter coefficient is close to the 10% significance level.

\textsuperscript{32}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.
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 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 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.

\footnote{In order to investigate the potential concerns related to the fact that the secondary schooling track results from a choice, and as such may be endogenous, we also ran our estimations without including the secondary schooling track variables. Our results, and in particular the estimates of the non-pecuniary factors, are overall robust to this specification change.}
<table>
<thead>
<tr>
<th>Variable</th>
<th>Unconstrained</th>
<th>Constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($\delta_0$)</td>
<td>-0.185 (0.174)</td>
<td>-0.026 (0.155)</td>
</tr>
<tr>
<td>Local average income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher education graduates</td>
<td>-0.026 (0.017)</td>
<td>-0.014* (0.008)</td>
</tr>
<tr>
<td>High school graduates</td>
<td>0.01 (0.012)</td>
<td>0</td>
</tr>
<tr>
<td>Secondary schooling track</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-0.288*** (0.087)</td>
<td>-0.142*** (0.054)</td>
</tr>
<tr>
<td>ES</td>
<td>-0.336*** (0.097)</td>
<td>-0.172*** (0.058)</td>
</tr>
<tr>
<td>S</td>
<td>-0.349*** (0.097)</td>
<td>-0.175*** (0.061)</td>
</tr>
<tr>
<td>Vocational</td>
<td>0.62** (0.248)</td>
<td>0.293* (0.164)</td>
</tr>
<tr>
<td>Technical</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Born abroad</td>
<td>-0.084** (0.033)</td>
<td>-0.031 (0.021)</td>
</tr>
<tr>
<td>Father born abroad</td>
<td>-0.034* (0.02)</td>
<td>-0.005 (0.011)</td>
</tr>
<tr>
<td>Mother born abroad</td>
<td>0.003 (0.014)</td>
<td>-0.009 (0.013)</td>
</tr>
<tr>
<td>Entering the labor market in 1998 (relative to 1992)</td>
<td>-0.272*** (0.084)</td>
<td>-0.12** (0.051)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.062*** (0.015)</td>
<td>-0.038** (0.009)</td>
</tr>
<tr>
<td>Father’s profession</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>-0.029 (0.02)</td>
<td>-0.023 (0.017)</td>
</tr>
<tr>
<td>Tradesman</td>
<td>-0.053*** (0.02)</td>
<td>-0.025** (0.011)</td>
</tr>
<tr>
<td>Executive</td>
<td>-0.105*** (0.034)</td>
<td>-0.054** (0.022)</td>
</tr>
<tr>
<td>Intermediate occupation</td>
<td>-0.071*** (0.025)</td>
<td>-0.035** (0.011)</td>
</tr>
<tr>
<td>Blue collar</td>
<td>0.000 (0.012)</td>
<td>-0.004 (0.008)</td>
</tr>
<tr>
<td>Other</td>
<td>-0.036** (0.015)</td>
<td>-0.023** (0.011)</td>
</tr>
<tr>
<td>White collar</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Mother’s profession</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>0.091** (0.039)</td>
<td>0.057 (0.037)</td>
</tr>
<tr>
<td>Tradesman</td>
<td>0.021 (0.019)</td>
<td>-0.003 (0.011)</td>
</tr>
<tr>
<td>Executive</td>
<td>-0.056*** (0.02)</td>
<td>-0.023* (0.014)</td>
</tr>
<tr>
<td>Intermediate occupation</td>
<td>-0.018 (0.013)</td>
<td>-0.019* (0.011)</td>
</tr>
<tr>
<td>Blue collar</td>
<td>0.076*** (0.027)</td>
<td>0.019* (0.01)</td>
</tr>
<tr>
<td>Other</td>
<td>0.012 (0.014)</td>
<td>-0.01 (0.007)</td>
</tr>
<tr>
<td>White collar</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Age in 6th grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\leq 10$</td>
<td>-0.103*** (0.038)</td>
<td>-0.047** (0.024)</td>
</tr>
<tr>
<td>11</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>$\geq 12$</td>
<td>0.108*** (0.041)</td>
<td>0.056** (0.026)</td>
</tr>
<tr>
<td>Paris region</td>
<td>-0.082*** (0.025)</td>
<td>-0.03** (0.012)</td>
</tr>
<tr>
<td>Vocational $\times$ ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entering the labor market in 1998</td>
<td>0.068** (0.029)</td>
<td>0.034 (0.024)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.02 (0.021)</td>
<td>0.003 (0.014)</td>
</tr>
<tr>
<td>Paris region</td>
<td>0.126*** (0.048)</td>
<td>0.059** (0.029)</td>
</tr>
</tbody>
</table>

Standard errors, presented in parentheses, were computed by bootstrap with 200 sample replicates. Significance 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 $u \approx 0.65$. The upper bound of the distribution can be used to compute a lower bound $E$ on the average return to higher education $E(Y_1 - Y_0)$. We obtain $E \approx 0.12$, 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.

As expected, the distribution of the *ex ante* returns is shifted towards the right for the subsample of higher education attendees, with a close to 10% probability of having a negative *ex ante* return, versus 28% for the whole sample. Hence, about 10% of the

---

$^34$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.

$^35$Indeed, an integration by part shows that

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

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

---
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 notably eight times larger than the 1.4 point decrease which is found to be 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).

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Ex ante return</th>
<th>Non-pecuniary factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>-0.069</td>
<td>-0.430</td>
</tr>
<tr>
<td>50%</td>
<td>0.133</td>
<td>-0.326</td>
</tr>
<tr>
<td>75%</td>
<td>0.267</td>
<td>-0.191</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Decile of G</th>
<th>Predicted attendance rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%, $G = -0.490$</td>
<td>0.952</td>
</tr>
<tr>
<td>90%, $G = 0.073$</td>
<td>0.630</td>
</tr>
</tbody>
</table>

5.5 Robustness checks

We address in the following two potential concerns about our results, namely the validity of our identification strategy and the robustness of our results to the assumptions made when computing the streams of earnings.\(^{36}\)

5.5.1 Validity of the identification 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 \textit{ex ante} returns 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 départements 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

\(^{36}\) 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 attempt 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 départements in productivity levels.\footnote{The Panel 1989 is a longitudinal dataset that follows 22,000 students entering 6th grade in 1989.} The estimates of the non-pecuniary factors
(see Panel 2, Table 7) as well as of the distribution of the \textit{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 \textit{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. Overall, our results are robust to alternative computations of $Y$.\footnote{We also estimate the streams of earnings where people are aware of their own annual increase $\rho_i$
of log-earnings, instead of just anticipating an average increase. We estimate $\rho_i$ by OLS and compute
the corresponding streams of earnings. The signs of $\gamma$ remain the same but no coefficient is significant
anymore. This can be explained by i) the importance of the errors on the estimated $\rho_i$ and ii) the fact
that the sample we can use in this case comprises only 9,364 individuals.}

6 Conclusion

This paper considers the identification of an extended Roy model, with a focus on the
non-pecuniary component of the selection equation. Our main theoretical contribution is
to prove that the identification of the covariates effects on potential earnings entails the
identification of this component as soon as one covariate is continuous. Using the detailed
structure of the selection equation is indeed sufficient to recover this non-pecuniary factor.
Notably, no exclusion restriction nor large support condition on the covariates 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 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).

Having identified the covariates effects on earnings and the non-pecuniary component of the sectoral choice, we also derive informative bounds on the distribution of the *ex ante* monetary returns, in the absence of exclusion restrictions. We propose a three-stage semi-parametric estimation procedure yielding root-n consistent and asymptotically normal estimators, the last stage allowing us 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 would be to build on our approach to conduct inference on the dependence between the sector-specific unobserved productivity terms \( \eta_0 \) and \( \eta_1 \). From an economic point of view, providing identification results on the joint distribution of these productivity components is especially worthwhile since it conveys information about the relative importance of general vs. specific human capital. This dependence has received much attention in competing risks models (see, e.g., Peterson, 1976, van den Berg, 1997, Abbring & van den Berg, 2003), but less so in the extensions of Roy models considered in the literature. A first approach to deal with this problem has been to specify a factor model (see, e.g., Carneiro et al., 2003), a second to use properties of analytic functions together with large support restrictions (see Fox & Gandhi, 2011).\(^{39}\)

We leave alternative analyses on this issue for further research.

\(^{39}\)Fox & Gandhi (2011) also show that without such large support restrictions, set identification is still possible.
7 Appendix A: proofs

Proposition 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,

$$E(\varepsilon_1|D = 1, X = x) = \frac{E(\eta_1 D|X = x)}{P(D = 1|X = x)}$$

$$= \frac{E (\eta_1 1\{\eta_1 \geq \psi_0(x) - \psi_1(x) + G(x)\})}{P(D = 1|X = x)}$$

(7.1)

Now let us show that almost surely,

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

(7.2)

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

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

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

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

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,

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

Suppose that there exists $\tilde{\psi}_1$ and $\tilde{h}$ such that

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

Then

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

By the measurably separation condition, this implies that $\tilde{\psi}_1$ and $\psi_1$ are almost surely equal up to a constant. This constant is identified by Assumption 2.2. Thus, $\psi_1$ is identified. $\psi_0$ can be recovered by the same argument.
Proposition 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 $k \in \{0, 1\}$,

$$P(D = k|X = x, Y_k = y) = P(D = k|X = x, \epsilon_k = y - \psi_k(x))$$

$$= P(\eta_k - \eta_{1-k} > \psi_{1-k}(x) - \psi_k(x) + G(x)|\eta_k + \nu_k = y - \psi_k(x)).$$

Thus, by Condition (iii) of Assumption 2.5,

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

for all $x$.

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

Theorem 2.3

First, note that

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

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

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

Besides, by definition of the supremum, there exists a sequence $(x_{n1})_{n \in \mathbb{N}}$ which tends to $\bar{x}_1$ and such that $x_{n1} \in A_{x_{-1}}$ 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 \( f_0(., \zeta_0) = f(., \zeta_0), q_0(., \zeta_0) = q(., \zeta_0) \) and \( r_0(.) = q_0(.) f_0(.) \). Consider the kernel estimators

\[
\hat{f}(u, \zeta) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left( \frac{u-X'_i \zeta}{h_n} \right)
\]

and \( \hat{r}(., \zeta) = \hat{q}(., \zeta) \times \hat{f}(., \zeta) \), where \( \hat{q}(., \zeta) \) is defined by Equation (3.6). Let us also define \( Z_i(\zeta) = 1\{X_i \in X\} h(X'_i \zeta) \) and, for any \( \mu = (r(., \zeta), \zeta, \tilde{\beta}_0, \tilde{\beta}_1) \), \( V_i(\mu) = D_i X'_i \zeta - \int_{u_0}^{X'_i \zeta} \frac{r(u)}{f(u)} \, du \).

We then let \( W_i(\mu) = (1, D_i, V_i(\mu))' \). Thus, \( \hat{W}_i = W_i(\hat{\mu}) \) and \( W_i = W_i(\mu_0) \), with \( \hat{\mu} = (\hat{r}(., \zeta), \hat{f}(., \zeta), \hat{\zeta}, \hat{\beta}_0, \hat{\beta}_1) \) and \( \mu_0 = (r_0, f_0, \zeta_0, \beta_0, \beta_1) \). Similarly, let \( \varepsilon_i(\mu) = Y_i - X'_i \left( D_i \hat{\beta}_1 + (1 - D_i) \hat{\beta}_0 \right) \).

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 \( E[g(A_i, \mu_0)] = 0 \) and

\[
\sum_{i=1}^{n} g(A_i, \hat{\theta}, \hat{\mu}) = 0.
\]

Thus, \( \hat{\theta} \) 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 \( \zeta \) 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

\[
G(A_i, \mu) = \xi_i \frac{\partial Z_i(\zeta_0)'}{\partial \zeta} + Z_i(\zeta_0) \left[ -X'_i (D_i \hat{\beta}_1 + (1 - D_i) \hat{\beta}_0) - \left( D_i X'_i \zeta \right. \right.
\]

\[
- q_0(U_i) X'_i \zeta - \int_{u_0}^{U_i} \frac{\partial q}{\partial \zeta}(u, \zeta_0)' \zeta + \frac{1}{f_0(u)} (r(u) - q_0(u) f(u)) \, du \left) \partial \alpha_0 \right].
\]

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

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g(A_i, \tilde{\mu}) - g(A_i, \mu_0) - G(A_i, \tilde{\mu} - \mu_0) \right] = o_P(1). \tag{7.3}
\]
For that purpose, we use the decomposition
\[ g(A_i, \hat{\mu}) - g(A_i, \mu_0) - G(A_i, \hat{\mu} - \mu_0) = R_{1i} + R_{2i} + R_{3i} + R_{4i} + R_{5i} \]
where, denoting by $h'(.)$ the vector of derivatives of $h(.)$ and $\bar{q} = \bar{r}/\bar{f}$, we let
\[ R_{1i} = \xi_i \mathbb{I}\{X_i \in \mathcal{X}\} \left(h(\hat{U}_i) - h(U_i) - (\hat{U}_i - U_i)h'(U_i)\right), \]
\[ R_{2i} = \alpha_0 Z_i(\zeta_0) \left[ \int_{U_i}^{\hat{U}_i} \bar{q}(u, \hat{\zeta}) du - q_0(U_i)(\hat{U}_i - U_i) \right], \]
\[ R_{3i} = \alpha_0 Z_i(\zeta_0) \int_{u_0}^{\hat{U}_i} \bar{q}(u, \hat{\zeta}) - \bar{q}(u) - \frac{\partial q}{\partial \zeta} (u, \zeta_0)(\hat{\zeta} - \zeta_0) du, \]
\[ R_{4i} = \alpha_0 Z_i(\zeta_0) \int_{u_0}^{\hat{U}_i} \bar{q}(u) - q_0(u) - \frac{1}{f_0(u)} \left( \bar{r}(u) - r_0(u) - q_0(u)(\bar{f}(u) - f_0(u)) \right) du, \]
\[ R_{5i} = [\varepsilon_i(\hat{\mu}) - \varepsilon_i(\mu_0) - (W_i(\hat{\mu}) - W_i(\mu_0))'\theta_0] \left[Z_i(\hat{\zeta}) - Z_i(\zeta_0)\right]. \]
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 $\|\cdot\|$ denotes the euclidian norm. Then, by the Cauchy-Schwarz inequality, $|\hat{U}_i - U_i| \leq C_0 |\hat{\zeta} - \zeta_0|$. Thus, by Assumptions 3.4 and 3.7,
\[ \sqrt{n} \max_{i=1,...,n} \left| h(\hat{U}_i) - h(U_i) - (\hat{U}_i - U_i)h'(U_i) \right| \leq \sqrt{n} M \max_{i=1,...,n} |\hat{U}_i - U_i|^2 \]
\[ \leq MC_0^2 \sqrt{n} |\hat{\zeta} - \zeta_0|^2 \]
\[ = o_P(1), \]
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, $\mathcal{S}_0 \subseteq \mathcal{S}$, where $\mathcal{S}$ denotes the support of $U$. Besides, by definition, $Z_i(\zeta_0) = Z_i(\zeta_0)\mathbb{I}\{U_i \in \mathcal{S}_0\}$. Moreover, for all $i$ such that $\hat{U}_i \in \mathcal{S}_0$, there exists, by the mean value theorem, $\tilde{U}_i = tU_i + (1 - t)\hat{U}_i$, with $t \in [0,1]$, such that $\int_{U_i}^{\hat{U}_i} q_0(u) du = q_0(\tilde{U}_i)(\tilde{U}_i - U_i)$. Thus, when $\hat{U}_i \in \mathcal{S}_0$,
\[ \|R_{2i}\| = \left\| \alpha_0 Z_i(\zeta_0)\mathbb{I}\{U_i \in \mathcal{S}_0\} \left\{ \int_{U_i}^{\hat{U}_i} \bar{q}(u, \hat{\zeta}) - q_0(u) \right\} du + \int_{U_i}^{\hat{U}_i} q_0(u) du - q_0(U_i)(\hat{U}_i - U_i) \right\| \]
\[ \leq C_1 |\tilde{U}_i - U_i| \left[ \sup_{u \in \mathcal{S}_0} |\bar{q}(u, \hat{\zeta}) - q_0(u)| + \max_{i: \tilde{U}_i \in \mathcal{S}_0} |q_0(\tilde{U}_i) - q_0(U_i)| \right] \]
\[ \leq C_0 C_1 \left\| \hat{\zeta} - \zeta_0 \right\| \left[ \sup_{u \in \mathcal{S}_0} |\bar{q}(u, \hat{\zeta}) - q_0(u)| + \max_{i: \tilde{U}_i \in \mathcal{S}_0} |q_0(\tilde{U}_i) - q_0(U_i)| \right], \]
38
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 $\hat{q}(., \hat{\zeta})$ and $q_0(.)$ are bounded above by 1, we have, when $\hat{U}_i \notin S_0$,

$$\|R_{2i}\| \leq 2C_0 C_1 \|\hat{\zeta} - \zeta_0\| 1\{U_i \in S_0\}.$$ 

Hence,

$$\|R_{2i}\| \leq C_0 C_1 \|\hat{\zeta} - \zeta_0\| \left(\sup_{u \in S_0} |\hat{q}(u, \hat{\zeta}) - q_0(u)| + \max_{i: \hat{U}_i \in S} |q_0(\hat{U}_i) - q_0(U_i)|\right) + 21 \{U_i \in S_0, \hat{U}_i \notin S_0\}.$$ 

(7.4)

By Assumption 3.4, $\sqrt{n} \|\hat{\zeta} - \zeta_0\| = O_P(1)$. Let us now show that the term into brackets in (7.4) is a $o_P(1)$. By Lemma 8.2, $\sup_{u \in S_0} |\hat{q}(u, \hat{\zeta}) - q_0(u)| = O_P(1)$. Now fix $\varepsilon > 0$. Because $q_0(.)$ is continuous by Assumption 2.3 and $S$ is compact, $q_0(.)$ is uniformly continuous on $S$. Thus, there exists $\delta > 0$ such that for all $(u, v) \in S^2$ satisfying $|u - v| \leq \delta$, we have $|q_0(u) - q_0(v)| \leq \varepsilon$. As a consequence,

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

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

$$\max_{i: \hat{U}_i \in S} |q_0(\hat{U}_i) - q_0(U_i)| = o_P(1).$$

It remains to show that

$$\frac{1}{n} \sum_{i=1}^{n} 1\{U_i \in S_0, \hat{U}_i \notin S_0\} = o_P(1).$$

(7.5)

For all $\delta > 0$, let $S_\delta = \{s \in S_0/\exists s' \notin S_0/|s - s'| < \delta\}$. Fix $\varepsilon > 0$ and let $K > 0$ be such that $P(U_i \in S_K) < \varepsilon/2$. For $n$ large enough, $P(C_0 \|\hat{\zeta} - \zeta_0\| > K) < \varepsilon/2$. Because $|U_i - \hat{U}_i| \leq C_0 \|\hat{\zeta} - \zeta_0\|$, we have, for $n$ large enough,

$$P \left( U_i \in S_0, \hat{U}_i \notin S_0 \right) \leq \frac{\varepsilon}{2} + P \left( U_i \in S_0, \hat{U}_i \notin S_0, C_0 \|\hat{\zeta} - \zeta_0\| \leq K \right) \leq \frac{\varepsilon}{2} + P(U_i \in S_K) \leq \varepsilon.$$

Because $\varepsilon$ was arbitrary, this proves that

$$E \left[ \frac{1}{n} \sum_{i=1}^{n} 1\{U_i \in S_0, \hat{U}_i \notin S_0\} \right] \rightarrow 0.$$
This establishes (7.5) since convergence in $L^1$ implies convergence in probability. As a result, $\sum_{i=1}^n R_{3i}/\sqrt{n} = o_P(1)$.

- $R_{3i}$: By the mean value theorem, there exists $\tilde{\zeta}_u$ in the segment between $\zeta_0$ and $\tilde{\zeta}$ such that
  \[\tilde{q}(u, \tilde{\zeta}) - \tilde{q}(u) = \frac{\partial \tilde{q}}{\partial \zeta}(u, \tilde{\zeta}_u)(\tilde{\zeta} - \zeta_0).\]

Because $U_i$ is bounded, there exists $C_2$ such that $|U_i - u_0| < C_2$. Thus,
\[
|R_{3i}| = \|\alpha_0 Z_i(\zeta_0)\| \left|\int_{u_0}^{U_i} \frac{1}{\tilde{f}(u) f_0(u)} \left[|\tilde{f}(u) - f_0(u)|^2 + |\tilde{r}(u) - r_0(u)|^2\right] du\right|
\leq C_1 C_2 \left\|\tilde{\zeta} - \zeta_0\right\| \sup_{u_0 S_0} \left\|\frac{\partial \tilde{q}}{\partial \zeta}(u, \tilde{\zeta}_u) - \frac{\partial q}{\partial \zeta}(u, \zeta_0)\right\|.
\]

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
\[
|R_{4i}| \leq C_1 \mathbb{1}\{U_i \in S_0\} \int_{u_0}^{U_i} \frac{1}{\tilde{f}(u) f_0(u)} \left[|\tilde{f}(u) - f_0(u)|^2 + |\tilde{r}(u) - r_0(u)|^2\right] du
\leq \frac{2C_1 C_2}{\inf_{u \in S_0} \tilde{f}(u) \inf_{u \in S_0} f_0(u)} \left[\left(\sup_{u \in S_0} |\tilde{f}(u) - f_0(u)|\right)^2 + \left(\sup_{u \in S_0} |\tilde{r}(u) - r_0(u)|\right)^2\right].
\]

Assumption 3.2 implies that the density of $U_i$ is positive in the interior of $S$. Thus, $\inf_{u \in S_0} f_0(u) > 0$. By uniform consistency of $\tilde{f}$ on $S_0$ (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 $\sup_{u \in S_0} |\tilde{f}(u) - f_0(u)| = o_P(n^{-1/4})$ and similarly for $\tilde{r}$. 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
\[
|\varepsilon_i(\tilde{\mu}) - \varepsilon_i(\mu_0) - (W_i(\tilde{\mu}) - W_i(\mu_0))' \theta_0| \mathbb{1}\{X_i \in \mathcal{X}\}
= \left|X_i'(D_i(\beta_1 - \hat{\beta}_1) + (1-D_i)(\beta_0 - \hat{\beta}_0)) + (D_i(U_i - \hat{U}_i) + \int_{U_i}^{U_i} \tilde{q}(u, \tilde{\zeta}) du
+ \int_{u_0}^{U_i} \left[\tilde{q}(u, \tilde{\zeta}) - q_0(u)\right] du\right) \alpha_0 \mathbb{1}\{X_i \in \mathcal{X}\}
\leq C_0 \left(\|\beta_1 - \beta_0\| + \|\beta_0 - \hat{\beta}_0\| + 2|\alpha_0|\|\tilde{\zeta} - \zeta_0\|\right) + C_2 |\alpha_0| \sup_{u \in S_0} |\tilde{q}(u, \tilde{\zeta}) - q_0(u)|.
\]

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 $\hat{U}_i$ and $U_i$ for all $i$. Thus, because $h'$ is continuous, there exists $C_3 > 0$ such that, with probability
approaching one,
\[ \|Z_i(\hat{\zeta}) - Z_i(\zeta_0)\| \leq C_3 \|\hat{\zeta} - \zeta_0\|. \]

Hence, with probability approaching one,
\[
\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{5i} \right| \leq \left[ C_0 C_3 \sqrt{n} \|\hat{\zeta} - \zeta_0\| \right] \left[ \|\hat{\beta}_1 - \beta_1\| + \|\hat{\beta}_0 - \beta_0\| + 2|\alpha_0| \|\hat{\zeta} - \zeta_0\| \right] + C_2 |\alpha_0| \sup_{u \in S_0} |\hat{g}(u, \hat{\zeta}) - q_0(u)|.
\]

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 $1/\sqrt{n} \sum_{i=1}^{n} G(A_i, \tilde{\mu} - \mu_0)$ can be linearized. Let $\kappa_0 = (\zeta_0, \beta_1, \beta_0)'$ and $\tilde{\kappa} = (\hat{\zeta}, \hat{\beta}_1, \hat{\beta}_0)'$. We have
\[ G(A_i, \tilde{\mu} - \mu_0) = P_i' (\tilde{\kappa} - \kappa_0) + \tilde{G}(A_i, \tilde{r}, \tilde{f}), \]
with $P_i = (P_{1i}, P_{2i}, P_{3i})'$ and
\[
\begin{align*}
P_{1i} &= \xi_i \frac{\partial Z_i}{\partial \zeta}(\zeta_0)' - \alpha_0 Z_i(\zeta_0) \left( D_i X_i' - q_0(U_i) X_i' - \int_{u_0}^{U_i} \frac{\partial q}{\partial \zeta}(u, \zeta_0) du \right) \\
P_{2i} &= -Z_i(\zeta_0) D_i X_i' \\
P_{3i} &= -Z_i(\zeta_0)(1 - D_i) X_i'
\end{align*}
\]
\[ \tilde{G}(A_i, \tilde{r}, \tilde{f}) = \alpha_0 Z_i(\zeta_0) \int_{u_0}^{U_i} (1/f_0(u))(\tilde{r}(u) - q_0(u)\tilde{f}(u)) du. \]

By the weak law of large numbers,
\[ \frac{1}{n} \sum_{i=1}^{n} P_i \xrightarrow{P} E[P]. \]
Moreover, by Assumptions 3.4 and 3.5,
\[ \sqrt{n}(\tilde{\kappa} - \kappa_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\chi_i, \chi_{1i}, \chi_{0i})' + o_P(1). \]

Thus,
\[
\left( \frac{1}{n} \sum_{i=1}^{n} P_i \right)' \sqrt{n}(\tilde{\kappa} - \kappa_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega_{1i} + o_P(1), \tag{7.7}
\]
where
\[ \Omega_{1i} = E[P_i' (\chi_i, \chi_{1i}, \chi_{0i})']. \tag{7.8}
\]
Thus, it suffices to focus on the nonparametric part of $G$, $\tilde{G}(A_i, \tilde{r}, \tilde{f})$. The main insight here is that $\tilde{G}$ 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

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

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

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

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

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

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

Now,

\[
Z_i(\hat{\zeta})W_i(\hat{\mu})' = Z_i(\zeta_0)W_i(\mu_0)' + Z_i(\hat{\zeta})(W_i(\hat{\mu}) - W_i(\mu_0))' + (Z_i(\zeta) - Z_i(\zeta_0))W_i(\mu_0)'.
\]

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

\[
\left\| W_i(\hat{\mu}) - W_i(\mu_0) \right\| \leq 2C_0 \left\| \hat{\zeta} - \zeta_0 \right\| + C_2 \sup_{u \in S_0} |\hat{q}(u, \hat{\zeta}) - q_0(u)|.
\]

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

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

Thus, by the weak law of large numbers,

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

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

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

This concludes the proof.
8 Appendix B: technical lemmas

Lemma 8.1 Suppose that Assumptions 2.3 and 3.2 hold. Then, for all \( u \in S \), the support of \( U \), \( \zeta \mapsto f(u, \zeta) \) and \( \zeta \mapsto r(u, \zeta) \), the density of \( X' \zeta \) and the derivative of \( u \mapsto E(D1\{X' \zeta \leq u\}) \) respectively, admit partial derivatives at \( \zeta_0 \) which satisfy:

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

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

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

\[
f(u, \zeta + t\delta_k) = \begin{cases} E \left[ f_{X_m|X_m} \left( \frac{u - X_m' \zeta_{m-1} - tX_k}{\zeta_m}, X_m \right) \right] & \text{if } k \neq m, \\
E \left[ f_{X_m|X_m} \left( \frac{u - X_m' \zeta_{m-1} + tX_k}{\zeta_m + t}, X_m \right) \right] & \text{if } k = m.
\end{cases}
\]

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

\[
F(u, \zeta + t\delta_k) = \begin{cases} E \left[ F_{X_m|X_m} \left( \frac{u - X_m' \zeta_{m-1} - tX_k}{\zeta_m}, X_m \right) \right] & \text{if } k \neq m, \\
E \left[ F_{X_m|X_m} \left( \frac{u - X_m' \zeta_{m-1} + tX_k}{\zeta_m + t}, X_m \right) \right] & \text{if } k = m.
\end{cases}
\]

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

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

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

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

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 \( G_0(u, \zeta) = E(D1\{X' \zeta \leq u\}) \) instead of \( F(u, \zeta) \). The partial derivatives of \( \zeta \mapsto G_0(u, \zeta) \) exist and satisfy

\[
\frac{\partial G_0}{\partial \zeta}(u, \zeta) = - E (DX|U = u) f_0(u) = -S_{\eta\Delta}(u + \delta_0) E (X|U = u) f_0(u).
\]
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 \( n h_n^6 \to \infty, n h_n^8 \to 0 \) and Assumptions 3.2 and 3.6 hold. Then, for all closed interval \( S' \) strictly included in the interior of \( S \) and for all \( \zeta_{u,n} \) such that \( \sup_{u \in S'} \|\zeta_{u,n} - \zeta_0\| = O_P(1/\sqrt{n}) \), we have,

\[
\sup_{u \in S'} |\hat{q}(u, \zeta_{u,n}) - q_0(u)| = o_P(1) \tag{8.3}
\]

\[
\sup_{u \in S'} \left\| \frac{\partial \hat{q}}{\partial \zeta} (u, \zeta_{u,n}) - \frac{\partial q}{\partial \zeta} (u, \zeta_0) \right\| = o_P(1) \tag{8.4}
\]

**Proof:** we first write

\[
\sup_{u \in S'} |\hat{q}(u, \zeta_{u,n}) - q_0(u)| \leq \sup_{u \in S'} |\hat{q}(u, \zeta_{u,n}) - \hat{q}(u, \zeta_0)| + \sup_{u \in S'} |\hat{q}(u, \zeta_0) - q_0(u)| \tag{8.5}
\]

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

\[
\sup_{u \in S'} |\hat{q}(u, \zeta_{u,n}) - \hat{q}(u, \zeta_0)| = \sup_{u \in S'} \left\| \frac{\hat{r}(u, \zeta_{u,n}) - \hat{r}(u, \zeta_0)}{\hat{f}(u, \zeta_0)} \right\|
\]

\[
\leq \sup_{u \in S'} \frac{1}{\hat{f}(u, \zeta_0)} \left[ |\hat{r}(u, \zeta_{u,n}) - \hat{r}(u, \zeta_0)| + |\hat{f}(u, \zeta_{u,n}) - \hat{f}(u, \zeta_0)| \right]
\]

\[
\leq \frac{1}{\inf_{u \in S'} \hat{f}(u, \zeta_0)} \left[ \sup_{u \in S'} |\hat{r}(u, \zeta_{u,n}) - \hat{r}(u, \zeta_0)| + \sup_{u \in S'} |\hat{f}(u, \zeta_{u,n}) - \hat{f}(u, \zeta_0)| \right]. \tag{8.6}
\]

Let us prove that

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

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

\[
\left| \hat{f}(u, \zeta_{u,n}) - \hat{f}(u, \zeta_0) \right| \leq \frac{1}{nh_n} \sum_{i=1}^n \left| K \left( \frac{u - X'_i \zeta_{u,n}}{h_n} \right) - K \left( \frac{u - X'_i \zeta_0}{h_n} \right) \right|
\]

\[
\leq \frac{C_4 C_0 \|\zeta_{u,n} - \zeta_0\|}{h_n^2}
\]

\[
\leq \frac{C_4 C_0 \sup_{u \in S'} \|\zeta_{u,n} - \zeta_0\|}{h_n^2} = O_p \left( \frac{1}{\sqrt{nh_n^2}} \right).
\]

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

\[
\inf_{u \in S'} \hat{f}(u, \zeta_0) \geq - \sup_{u \in S'} \left| \hat{f}(u, \zeta_{u,n}) - \hat{f}(u, \zeta_0) \right| + \inf_{u \in S'} f_0(u),
\]

...
and because \( \inf_{u \in S'} f_0(u) > 0 \) by Assumption 3.2, we also have
\[
\frac{1}{\inf_{u \in S'} \hat{f}(u, \zeta_0)} = O_p(1).
\]
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
\[
\sup_{u \in S'} |\hat{f}(u, \zeta_0) - f_0(u)| = o_p(1)
\]
and similarly for \( \hat{r}(\cdot, \zeta_0) \). This establishes (8.3).

Now, let us turn to (8.4). We use the same decomposition as (8.5). First, let us establish that
\[
\sup_{u \in S'} \left| \frac{\partial \hat{q}}{\partial \zeta}(u, \zeta_0) - \frac{\partial q}{\partial \zeta}(u, \zeta_0) \right| = o_P(1) \tag{8.8}
\]
We have
\[
\frac{\partial \hat{q}}{\partial \zeta}(u, \zeta_0) = \frac{1}{f(u, \zeta_0)} \left[ \frac{\partial \hat{r}}{\partial \zeta}(u, \zeta_0) - \hat{q}(u, \zeta_0) \frac{\partial \hat{f}}{\partial \zeta}(u, \zeta_0) \right].
\]
and similarly for \( \partial q / \partial \zeta(u, \zeta_0) \). Thus,
\[
\frac{\partial \hat{q}}{\partial \zeta}(u, \zeta_0) - \frac{\partial q}{\partial \zeta}(u, \zeta_0) = \frac{1}{f(u, \zeta_0)} \left\{ \left[ \frac{\partial \hat{r}}{\partial \zeta}(u, \zeta_0) - \frac{\partial r}{\partial \zeta}(u, \zeta_0) \right] - \frac{\partial r / \partial \zeta(u, \zeta_0)}{f_0(u)} \left[ \hat{f}(u, \zeta_0) - f_0(u) \right] \right\}
- \frac{\hat{q}(u, \zeta_0)}{f(u, \zeta_0)} \left[ \frac{\partial \hat{f}}{\partial \zeta}(u, \zeta_0) - \frac{\partial f}{\partial \zeta}(u, \zeta_0) \right] - \frac{\partial f / \partial \zeta(u, \zeta_0)}{f_0(u)} \left( \hat{f}(u, \zeta_0) - f_0(u) \right) - \frac{\partial f / \partial \zeta(u, \zeta_0)}{f_0(u)} (\hat{q}(u, \zeta_0) - q_0(u)).
\]
By what precedes, \( \inf_{u \in S'} \hat{f}(u, \zeta_0) \) tends in probability to \( \inf_{u \in S'} f_0(u) > 0 \), while \( \sup_{u \in S'} |\hat{f}(u, \zeta_0) - f_0(u)| = o_P(1) \). Besides, \( \hat{q}(\cdot, \zeta_0) \) is bounded by 1 and by Lemma 8.1, \( \partial f / \partial \zeta(\cdot, \zeta_0) \) is continuous on the compact set \( S \) and thus is bounded on this set. Thus, it suffices to prove that
\[
\sup_{u \in S'} \left| \frac{\partial \hat{f}}{\partial \zeta}(u, \zeta_0) - \frac{\partial f}{\partial \zeta}(u, \zeta_0) \right| = o_P(1) \tag{8.9}
\]
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

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

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

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

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

$$\left\| \frac{\partial \hat{f}}{\partial \zeta}(u, \zeta_{u,n}) - \frac{\partial \hat{f}}{\partial \zeta}(u, \zeta_0) \right\| \leq \frac{1}{nh_n^2} \sum_{i=1}^{n} \left\| X_i \right\| \left| K'(\frac{u - X'_i\zeta_{u,n}}{h_n}) - K'(\frac{u - X'_i\zeta_0}{h_n}) \right|$$

$$\leq C_5C_6^2 \frac{\|\zeta_{u,n} - \zeta_0\|}{h_n^3} = O_p \left( \frac{1}{\sqrt{n}h_n^3} \right).$$

This proves (8.10) since $nh_n^6 \to \infty$. The same reasoning applies to $\hat{r}$. The result follows.
# Appendix C: supplementary tables and figures

## Variables

<table>
<thead>
<tr>
<th>Local average income</th>
<th>(\zeta)</th>
<th>(\beta_0)</th>
<th>(\beta_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher education graduates</td>
<td>1.541*** (0.087)</td>
<td>0</td>
<td>0.019*** (0.004)</td>
</tr>
<tr>
<td>High school graduates</td>
<td>-1 (0)</td>
<td>0.022*** (0.004)</td>
<td>0</td>
</tr>
</tbody>
</table>

### 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) |

### Born abroad

| Born abroad | 1.727*** (0.46) | -0.006 (0.017) | 0.00 (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)

| Entering the labor market in 1998 (relative to 1992) | 9.133*** (0.447) | 0.097*** (0.035) | 0.173*** (0.024) |

### Male

| 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

<table>
<thead>
<tr>
<th>Age in 6th grade</th>
<th>(\zeta)</th>
<th>(\beta_0)</th>
<th>(\beta_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\leq 10)</td>
<td>3.825*** (0.465)</td>
<td>-0.028 (0.017)</td>
<td>0.007 (0.01)</td>
</tr>
<tr>
<td>11</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>(\geq 12)</td>
<td>-5.07*** (0.425)</td>
<td>0.035* (0.018)</td>
<td>-0.019 (0.013)</td>
</tr>
</tbody>
</table>

### Paris region

| Paris region | 1.181*** (0.453) | 0.003 (0.012) | -0.002 (0.004) |

### Vocational × ...

| 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. Significance levels: *** (1%), ** (5%) and * (10%).
<table>
<thead>
<tr>
<th>Variable</th>
<th>Panel 1</th>
<th>Panel 2</th>
<th>Panel 3</th>
<th>Panel 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($\delta_0$)</td>
<td>-0.016 (0.171)</td>
<td>0.006 (0.175)</td>
<td>-0.028 (0.164)</td>
<td>-0.024 (0.155)</td>
</tr>
<tr>
<td>Local average income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher education graduates</td>
<td>-0.01 (0.007)</td>
<td>-0.013* (0.008)</td>
<td>-0.01 (0.008)</td>
<td>-0.014* (0.008)</td>
</tr>
<tr>
<td>Local rate of honours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary schooling track</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-0.128*** (0.046)</td>
<td>-0.132*** (0.049)</td>
<td>-0.117** (0.059)</td>
<td>-0.142*** (0.054)</td>
</tr>
<tr>
<td>ES</td>
<td>-0.154*** (0.05)</td>
<td>-0.162*** (0.052)</td>
<td>-0.15** (0.063)</td>
<td>-0.172*** (0.058)</td>
</tr>
<tr>
<td>S</td>
<td>-0.146*** (0.051)</td>
<td>-0.164*** (0.054)</td>
<td>-0.135** (0.066)</td>
<td>-0.175*** (0.061)</td>
</tr>
<tr>
<td>Vocational</td>
<td>0.227 (0.226)</td>
<td>0.351** (0.173)</td>
<td>0.251 (0.175)</td>
<td>0.293* (0.165)</td>
</tr>
<tr>
<td>Born abroad</td>
<td>-0.02 (0.02)</td>
<td>-0.032 (0.02)</td>
<td>-0.03 (0.022)</td>
<td>-0.032 (0.021)</td>
</tr>
<tr>
<td>Father born abroad</td>
<td>0 (0.01)</td>
<td>-0.005 (0.011)</td>
<td>-0.006 (0.012)</td>
<td>-0.005 (0.011)</td>
</tr>
<tr>
<td>Mother born abroad</td>
<td>-0.011 (0.011)</td>
<td>-0.006 (0.012)</td>
<td>-0.009 (0.014)</td>
<td>-0.009 (0.013)</td>
</tr>
<tr>
<td>Entering the labor market in 1998 (relative to 1992)</td>
<td>-0.094*** (0.034)</td>
<td>-0.106** (0.045)</td>
<td>-0.113** (0.055)</td>
<td>-0.12** (0.051)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.061*** (0.012)</td>
<td>-0.043*** (0.008)</td>
<td>-0.044*** (0.009)</td>
<td>-0.038*** (0.009)</td>
</tr>
<tr>
<td>Father’s profession</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>-0.02 (0.016)</td>
<td>-0.022 (0.016)</td>
<td>-0.018 (0.018)</td>
<td>-0.023 (0.017)</td>
</tr>
<tr>
<td>Tradesman</td>
<td>-0.021** (0.009)</td>
<td>-0.026** (0.013)</td>
<td>-0.02* (0.012)</td>
<td>-0.025** (0.011)</td>
</tr>
<tr>
<td>Executive</td>
<td>-0.051** (0.023)</td>
<td>-0.053** (0.022)</td>
<td>-0.043* (0.024)</td>
<td>-0.055** (0.022)</td>
</tr>
<tr>
<td>Intermediate occupation</td>
<td>-0.034** (0.014)</td>
<td>-0.04*** (0.015)</td>
<td>-0.03** (0.013)</td>
<td>-0.035*** (0.012)</td>
</tr>
<tr>
<td>Blue collar</td>
<td>-0.009 (0.007)</td>
<td>-0.007 (0.007)</td>
<td>-0.005 (0.009)</td>
<td>-0.004 (0.008)</td>
</tr>
<tr>
<td>Other</td>
<td>-0.016 (0.011)</td>
<td>-0.018 (0.011)</td>
<td>-0.021* (0.012)</td>
<td>-0.023** (0.011)</td>
</tr>
<tr>
<td>Mother’s profession</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>0.049 (0.034)</td>
<td>0.045 (0.03)</td>
<td>0.049 (0.039)</td>
<td>0.057 (0.037)</td>
</tr>
<tr>
<td>Tradesman</td>
<td>-0.008 (0.01)</td>
<td>0.002 (0.012)</td>
<td>-0.006 (0.012)</td>
<td>-0.003 (0.011)</td>
</tr>
<tr>
<td>Executive</td>
<td>-0.017 (0.012)</td>
<td>-0.018 (0.012)</td>
<td>-0.017 (0.015)</td>
<td>-0.023* (0.014)</td>
</tr>
<tr>
<td>Intermediate occupation</td>
<td>-0.018 (0.012)</td>
<td>-0.017 (0.011)</td>
<td>-0.019 (0.012)</td>
<td>-0.019* (0.011)</td>
</tr>
<tr>
<td>Blue collar</td>
<td>0.011 (0.007)</td>
<td>0.017* (0.01)</td>
<td>0.016 (0.01)</td>
<td>0.019* (0.01)</td>
</tr>
<tr>
<td>Other</td>
<td>-0.008 (0.007)</td>
<td>-0.01 (0.006)</td>
<td>-0.009 (0.007)</td>
<td>-0.01 (0.007)</td>
</tr>
<tr>
<td>Age in 6th grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 10</td>
<td>-0.037* (0.021)</td>
<td>-0.039** (0.019)</td>
<td>-0.033 (0.025)</td>
<td>-0.047** (0.024)</td>
</tr>
<tr>
<td>≥ 12</td>
<td>0.06* (0.036)</td>
<td>0.05** (0.024)</td>
<td>0.048* (0.028)</td>
<td>0.056** (0.026)</td>
</tr>
<tr>
<td>Paris region</td>
<td>-0.024* (0.012)</td>
<td>-0.023* (0.012)</td>
<td>-0.018 (0.014)</td>
<td>-0.03** (0.012)</td>
</tr>
<tr>
<td>Vocational × ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entering the labor market in 1998</td>
<td>0.019 (0.023)</td>
<td>0.04* (0.023)</td>
<td>0.02 (0.026)</td>
<td>0.034 (0.024)</td>
</tr>
<tr>
<td>Male</td>
<td>0.019 (0.016)</td>
<td>0.004 (0.015)</td>
<td>0.008 (0.015)</td>
<td>0.003 (0.014)</td>
</tr>
<tr>
<td>Paris region</td>
<td>0.045* (0.025)</td>
<td>0.052* (0.028)</td>
<td>0.038 (0.032)</td>
<td>0.059* (0.029)</td>
</tr>
</tbody>
</table>

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. Significance 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


51


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


