Inference on a Generalized Roy Model with Exclusion Restrictions

Xavier d’Haultfoeuille * Arnaud Maurel †

14th May 2009
Preliminary version

Abstract

This paper considers nonparametric identification and estimation of an extended Roy model, which is obtained by including covariates as well as a non-pecuniary component in the seminal Roy’s model (1951) of occupational choice. This framework is well suited to various economic contexts, including educational and sectoral choices, or labor supply decisions. We show that the effects of covariates on earnings are identified through exclusion restrictions or conditions at infinity. We also study identification of the non-pecuniary component when at least one variable affects the selection probability only through potential earnings, that is to say under the opposite of the usual exclusion restrictions used to identify switching regressions models. Point identification is obtained if such variables are continuous, while bounds are obtained otherwise. We propose a three-stages semiparametric estimation procedure for this model, which yield root-n consistent and asymptotically normal estimators. We also provide finite sample properties relying on Monte Carlo simulations. We finally apply our results to the educational context, by providing new evidence on the influence of non-pecuniary factors on the decision to attend higher education.

JEL Classification: C14, C35 and J24

Keywords: Roy model, nonparametric identification, exclusion restrictions and semiparametric estimation

*CREST-INSEE and Université Paris I-Panthéon-Sorbonne. E-mail address: xavier.dhaultfoeuille@ensae.fr.
†CREST-ENSAE, PSE(Paris) and IZA(Bonn). E-mail address: maurel@ensae.fr.
1 Introduction

This paper considers nonparametric identification and estimation of an extended Roy model, which is obtained by including covariates as well as a non-pecuniary component in the seminal Roy (1951) model of occupational choice. The Roy model has been widely applied in labor economics to consider the choice of which market (or sector) to enter. Among others, Roy’s framework has been applied to analyze female labor force participation (Heckman (1974)), the choice between union and nonunion status (Lee (1978), Robinson & Tomes (1984)) as well as between public and private sector (Dustmann & Soest (1998)), the choice of educational level (Willis & Rosen (1979)), migration decisions (Borjas (1987)), training program participation (Ashenfelter & Card (1985), Ham & LaLonde (1996)) and occupational choice (Dolton et al. (1989)).

Unlike the original Roy’s model (1951), which assumes that individuals pursue their comparative advantage by self-selecting themselves into the occupation yielding highest earnings, the framework we consider in the paper also allows to account for non-pecuniary aspects of occupational choice decisions. Including a non-pecuniary factor in the utility associated with each choice alternative is necessary in many economic contexts, under which the occupational choice cannot be properly described by considering the agents as solely maximizing their income. For instance, in the context of schooling choice, it is most often assumed that individuals consider not only the investment value, which is related to the wage returns, but also the non-pecuniary consumption value of each alternative, which is related to the tastes for each schooling alternative. Non-pecuniary aspects also matter in the context of occupational choice (e.g. stress, injury risk associated with each job) as well as migration decisions (e.g. psychic cost of migration). Furthermore, this extended Roy framework can also be applied to model choices within a risky environment, with the “non-pecuniary” component being related for instance to the probability of passing an entry examination for a job in the public sector.

This paper contributes to the extensive literature on the identification of Roy models, originated by Heckman & Honore (1990). Seminal papers by Cox(1959,1962) and Tsiatis (1975) provide nonidentifiability results for the closely related competing risks models. They show that, for any joint distribution of the latent failure times, there exists a joint distribution with independent failure times which generates the same distribution of the identified minimum. Adapting Cox (1962) and Tsiatis (1975), Heckman & Honore (1990) show that, when relaxing Roy’s log normality assumption, any-cross section wage distribution can be rationalized by a standard Roy model with two independent skills. This well known non-
identifiability result has led number of applied papers estimating Roy models in the context of occupational choices to be conducted within a quite restrictive independence paradigm. However, Heckman & Honore (1990) also show that including in the Roy model regressors with large support and imposing exclusion restrictions allows to overcome the preceding nonidentifiability result. Under this restrictive large support assumption, the extended Roy model with covariates is identified.¹ This result was extended, in the competing risks framework, by Abbring & van den Berg (2003) and Lee (2006) to the mixed proportional hazards and generalized accelerated failure time models respectively. They show that in these models, neither exclusion restrictions nor any large support assumptions are necessary to achieve identification. In our paper, we also give identification results for the generalized Roy model with non-pecuniary component without any large support condition. However, due to the extra complexity arising through the non-pecuniary component, we impose exclusion restrictions to yield identification.

On a related ground, a recent paper by Bayer et al. (2008) also consider the identification and the estimation of a generalized Roy model including a non-pecuniary factor,² without assuming the availability of regressors. Nevertheless, their identification results are obtained under two different sets of assumptions. First, they show that assuming that the distribution of pecuniary returns has a finite lower bound allows to identify the non-pecuniary factors associated with each choice alternative and the unconditional wage distributions. This identifying assumption can be restrictive, in particular when using log wages in utility functions, as for instance in Willis & Rosen (1979). In a paper addressing inference on randomly censored regression models, which include the semiparametric Roy model as a special case, Khan & Tamer (2009) also obtain point identification under a support condition implying in particular that the upper bound of the wage distribution is finite. Bayer et al. (2008) alternatively prove identification under the assumption of independence between alternative-specific wages and the exclusion restriction that a variable affects the nonpecuniary valuation of each choice alternative but not the wage distributions.³ The independence assumption, however, is restrictive, and, as mentioned before, much of the literature considering identification of Roy and competing risks models has produced alternative identification results obtained by relaxing this stringent assumption. The identification results we derive in our paper does not rely neither on support assumption nor on the independence assumption.

¹Under this condition, both the covariates effects as well as the joint distribution of sector-specific skills are nonparametrically identified.
²They also extend the pure Roy model by relying on a multinomial choice setting.
³Bayer et al. (2008) refer to this exclusion restriction as the Commonality assumption.
In the paper, we first prove identification of covariates effects on sector-specific earnings. Under an assumption of *moderate dependence* between the unobserved productivity terms, identification is achieved at infinity, in the same spirit as Heckman & Honore (1990). Nevertheless, unlike them, this result is obtained without relying on large support assumption nor on exclusion restrictions. We also show that exclusion restrictions, both between sector-specific regressors as well as between the non-pecuniary component of utility and these sector-specific regressors, secure identification of covariates effects without the need to rely on an argument at the limit. Then we study identification of the non-pecuniary component under the condition that at least one regressor affects the selection probability only through sector-specific potential earnings. This is the opposite of the kind of exclusion restrictions which are usually imposed to identify switching regressions or selection models. For instance, in the context of college attendance decision, local economic conditions are natural candidates for these exclusion restrictions. Under this assumption, we provide bounds on the non-pecuniary components. We show that the length of the identifying interval depends on the support of these instruments. In particular, point identification is achieved in general when at least one instrument is continuous.

We also propose a three-stages semiparametric estimation procedure for the model when the effects of the covariates are linear. The first two stages enable to estimate the effects of the covariates on earnings and correspond to Newey (2008)’s method for estimating semiparametric selection models. The third stage, which is devoted to the non-pecuniary component, is rather simple as it amounts to estimate an instrumental linear model. The only difficulty lies in estimating the dependent variable of this linear model, as it involves both the first steps estimators and a nonparametric nuisance parameter. We show that the corresponding estimator is root-n consistent and asymptotically normal. Monte Carlo simulations indicate that despite its multiple steps, the estimators perform relatively well in finite samples. Eventually, we conclude this paper by applying our results to the context of college attendance decision. In line with Carneiro et al. (2003), we provide new evidence on the impact of non-pecuniary factors on the decision to attend college, relying on French *Generation* dataset.

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

---

4Such an identification at infinity typically leads to slow rates of convergence, as for instance in Fermanian (2003).

5See also d’Haultfoeuille (2008) for a detailed analysis of the identifying power of such exclusion restrictions in the context of endogenous selection.

6Such instruments were used by Carneiro et al. (2003) to identify an extension of the Willis and Rosen’s(1979) model of demand for college attendance.
model with a non-pecuniary component which is considered throughout the paper and exhibits different economic contexts fitting in this framework. Section 3 gives identification results for the covariates effects on earnings and for the non-pecuniary component. Section 4 develops a semiparametric estimation procedure for the extended Roy model, and proves root-n consistency and asymptotic normality of the proposed estimators. Section 5 studies finite-sample performances of the estimators. Section 6 applies the preceding estimators to recover a semiparametric estimate of the impact of non-pecuniary factors on the college attendance decision. The proofs of our results are deferred to the appendix.

2 The model and examples

In this section, we first present an extension of the simple Roy model with covariates which accounts for non-pecuniary factors in the self-selection process. Then, we exhibit several toy models that lead to this extended Roy framework. In all of these situations, the non-pecuniary component turns out to have an interesting interpretation, so that recovering this factor can be worthwhile in various economic contexts.

2.1 A Roy model with non-pecuniary factors

In this paper, we provide identification and estimation results for an extension of the Roy model which is obtained by including covariates as well non-pecuniary factors in the seminal Roy’s model (1951) of occupational choice. This extension is presented in the following. Suppose that there are two sectors 0 and 1 in the economy, and let $Y_i$, $i \in \{0, 1\}$, denote the individual’s potential earnings in sector $i$. Sector-specific earnings $Y_i \in [0, +\infty[$ are assumed to depend on observed covariates $X$ and an unobserved sector-specific productivity term $\varepsilon_i$ in the following way:

$$Y_i = \exp (\psi_i(X) + \varepsilon_i) \quad (2.1)$$

It is noteworthy that, unlike in 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 utility, with the utility in sector $i$ writing as $U_i = G_i(Z) + \ln Y_i$. Hence, utility from choosing sector $i$ is assumed to be given by the sum of sector-specific log-earnings $\ln Y_i$ and the non-pecuniary component associated with sector $i$, $G_i(Z)$, which is supposed to depend on observed covariates $Z$. Thus, along with

\footnote{For the sake of simplicity and in the absence of ambiguity, individual subscripts are omitted until Section 4.}
the covariates $X$ and $Z$, we observe the chosen sector, defining $G \equiv G_0 - G_1$:

$$D = \mathbb{1}\{\ln Y_1 > G(Z) + \ln Y_0\} = \mathbb{1}\{\varepsilon_1 > G(Z) + \psi_0(X) - \psi_1(X) + \varepsilon_0\}$$

Besides, we also observe earnings in the chosen sector, which are expressed as:

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

In the following, we set up several simple discrete choice frameworks which are consistent with this generalized Roy model. The four examples presented below aim at shedding light on the practical relevance, in various economic contexts, of providing inference results for the Roy model with non-pecuniary factors.

2.2 Binary decisions with non-pecuniary aspects

2.2.1 Example 1: College attendance decision and consumption value of schooling

We consider the decision to attend college once graduating from high school. After completing secondary education, individuals are assumed to decide either to enter directly the labor market with a high school degree ($c = 0$) or to attend college ($c = 1$). They are supposed to make their decision ($c^*$) by comparing the expected streams of income related to each alternative.\(^8\)

When entering the labor market, individuals receive at each period $t$ the labor income $Y_{c,t}$ which is supposed to follow:

$$\ln Y_{c,t} = \rho_c + \ln Y_{c,t-1} + \varepsilon_{c,t}$$

Where $\rho_c$ denotes the degree $c$-specific return to experience and $\varepsilon_{c,t}$ is a degree $c$-specific unobserved individual productivity term which is assumed to be independently and identically distributed over time, with mean zero. We further assume that the initial labor income $Y_{c,0}$ is given by:

$$\ln Y_{c,0} = u_c + \varepsilon_{c,0}$$

Where $u_c \equiv \psi_c(X) + \eta_c$ is an individual term depending on the schooling choice $c$. $\psi_c(\cdot)$ are unknown functions of observed individual covariates $X$, while $\eta_c$ is a productivity random\(^8\) Willis & Rosen (1979) were the first to apply the Roy framework to model the decision to attend college.

Recently, Carneiro et al. (2003) estimate an extension of Willis and Rosen’s model including uncertainty and non-pecuniary factors. Note that this framework is suited to model a broad range of schooling as well occupational choices, including also for instance the choice between attending a vocational or a general college.
term which is supposed to be known by the individual at the time of her decision but unobserved by the econometrician.  

The indirect utility $V_c$ associated with each schooling decision $c$ is supposed to be given by, denoting by $\beta$ the discount factor, by $T$ the date of retirement and by $G_1(Z)$ the consumption value of schooling \(^9\) which depends on observed individual covariates $Z$:  

$$V_c = \sum_{t=0}^{T} \beta^t E(\ln Y_{c,t}|X, u_c) + G_c(Z)$$

Which can be rewritten as:

$$V_c = \left( \sum_{t=0}^{T} t\beta^t \right) \rho_c + \left( \sum_{t=0}^{T} \beta^t \right) (\psi_c(X) + \eta_c) + G_c(Z)$$

After graduating from high school, the individual makes the decision $c$ which yields the highest value $V_c$. Hence, the individual will choose to attend college ($c = 1$) if and only if:

$$\eta_1 > \frac{G_1(Z)}{\sum_{t=0}^{T} \beta^t} + \frac{\sum_{t=0}^{T} t\beta^t (\rho_0 - \rho_1)}{\sum_{t=0}^{T} \beta^t} + (\psi_1(X) - \psi_0(X)) + \eta_0$$

This example falls into the generalized Roy framework considered above.  

\(^{11}\) Access to longitudinal data allows us to recover the parameters $(\rho_0, \rho_1)$, and thus in this context, after normalizing $\beta$, recovering the non-pecuniary factor allows to estimate the consumption value of schooling $G_1(\cdot)$.

### 2.2.2 Example 2: Labor supply with non-pecuniary job attributes

The generalized Roy model considered throughout the paper can also be used to model labor supply decisions in the presence of non-pecuniary job attributes such as type of work as well as work hardness. Let us consider a simple model of labor supply in which each individual has to choose between two jobs. Each job is supposed to be characterized by the standard wage rate $w$ and hours of work $h$ variables, as well as other “qualitative” attributes. Denoting by $U(C, h, j)$ the utility function of the individual, where $C$ denotes her

---

\(^9\)We also assume independence between $\eta_c$ and $(\varepsilon_{c,t})_t$.

\(^{10}\)As opposed to the investment value of schooling, which corresponds in this example to the expected discounted flow of future log-earnings. Since we can only identify relative non-pecuniary components, we let $G_0 = 0$ without loss of generality.

\(^{11}\)Nevertheless, unlike the preceding Roy model, self-selection is only driven by $(\eta_0, \eta_1)$ and not by the complete unobserved factors $(\eta_0 + \varepsilon_{0,0}, \eta_1 + \varepsilon_{1,0})$. This is actually innocuous, since one can easily show that our identification scheme also holds with respect to this slightly different selection process.

\(^{12}\)Indeed, we can identify from the data $E(\ln Y_{c,0}|D = c) = E(u_c|D = c)$, and thus $\rho_c$ parameters are obtained by forming the differences: $E(\ln Y_{c,1}|D = c) - E(\ln Y_{c,0}|D = c) = \rho_c$. 

7
consumption and \( j \in \{0,1\} \) refers to each job \( j \), we assume the following multiplicatively separable specification, where \( G_j \) and \( \phi(.) \) are positive:\(^{13}\)

\[
U(C,h,j) = G_j \phi(C_j,h_j)
\]

Hence, assuming further (consistently with Altonji & Paxson (1988)) that hours of work are fixed with \( h_1 = h_2 = h \), we have (denoting by the \( D \) the dummy which is equal to one if job 1 is chosen, and zero otherwise):

\[
D = 1\{\phi(C_1,h) > \frac{G_0}{G_1} \phi(C_0,h)\}
\]

Now the budget constraint writes \( C_j = w_j h \). Hence, letting \( \psi(w) = \phi(wh,h) \), we finally have:

\[
D = 1\{\ln \psi(w_1) > \ln G_0 - \ln G_1 + \ln \psi(w_0)\}
\]

Which corresponds to the generalized Roy model presented above, with a non-pecuniary component equal to \( \ln G_0 - \ln G_1 \). In this context, recovering this term allows to estimate the relative valuation of non-pecuniary job attributes.

### 2.3 Binary decisions under risk

The generalized Roy model is also useful for modeling situations in which an individual has to make a choice between a risky and a certain prospect. Consider for instance the sectoral choice decision between private and public sector. The individual is assumed to decide either to work in the private sector (alternative A) or to take a public sector entrance exam in order to become a civil servant (alternative B).\(^{14}\) The two alternatives are described below:

- **Alternative A**: once entering the private sector labor market, the individual receives a stream of income denoted by \( Y_0 \).

- **Alternative B**: this alternative is risky in the sense that examination success is \textit{ex ante} uncertain to the individual. More precisely, we suppose that she will pass the

\(^{13}\)See, for instance, Dagsvik & Strøm (2006) for a similar specification on labor supply behavior in the presence of non-pecuniary job attributes.

\(^{14}\)Examples of related frameworks examining the determinants of queues for public sector jobs can be found in Krueger (1988) and Heywood & Mohanty (1994, 1995). The assumption according to which potential wages are perfectly observed at the moment of the choice is especially sensible in this context since public wages are regulated by the government at a certain rate which is easily observed.
public sector entrance exam with a probability $\pi$, and fail with a probability $1 - \pi$. Individuals entering the public sector are assumed to be endowed with a stream of income $Y_1 > Y_0$, while those who failed to the entrance exam enter the private sector and receive the stream of income $\alpha Y_0$ (with $\alpha < 1$).\(^{15}\)

Denoting by $D$ the dummy which is equal to one if alternative B (public sector) is chosen and zero otherwise, and assuming that the individual is a risk-neutral utility maximizer, we have:

$$D = 1\{\pi Y_1 + (1 - \pi)\alpha Y_0 > Y_0\}$$

Which can be rewritten as:

$$D = 1\{\ln Y_1 > \ln\left(\frac{1 - \alpha(1 - \pi)}{\pi}\right) + \ln Y_0\}$$

Supposing that the opportunity cost of exam preparation, and thus $\alpha$, are observed, recovering the term $\ln\left(\frac{1 - \alpha(1 - \pi)}{\pi}\right)$ will allow to obtain the completion probability $\pi$ which is taken into account by the individual when deciding which sector to enter.\(^{16}\) Thus, estimating this factor in this context is interesting since it will allow to compare the subjective completion probabilities $\pi$ with the observed ones.\(^{17}\)

Finally, aside from applications in labor and education economics, a similar framework can be applied to study issues in other fields such as empirical industrial organization.\(^{18}\)

### 3 Identification

In this section, we discuss the identifiability of functions $(\psi_0, \psi_1, G)$ in the Roy model with non-pecuniary component. The identification issue amounts to determine the set of

\(^{15}\)In this case, the stream of income is lower than when entering the private sector directly without taking the public sector entrance exam because of the costs, both direct and indirect, of preparing the exam. In order to express this stream of income as a constant share of $Y_0$, we further need to assume that:

i) the direct cost is neglectable relative to foregone earnings during exam preparation

ii) the returns to experience are constant across individuals.

\(^{16}\)Note that the factor $\ln\left(\frac{1 - \alpha(1 - \pi)}{\pi}\right)$ can be interpreted as a non-pecuniary component of the sectoral choice decision in the sense that it does not only depend on the relative monetary returns. It also depends on the probability $\pi$ of success to the public sector entrance exam.

\(^{17}\)The same framework can also be used in order to model schooling decisions under risk and to recover the subjective beliefs concerning future academic outcomes. On a related ground, recent papers by Nicholson (2002) and Belzil (2007) estimate schooling decisions models in which students hold subjective beliefs about their future academic outcomes.

\(^{18}\)Indeed, one can apply this framework in order to model e.g. the decision for a firm to engage in collusion, which is risky in the sense that it faces a financial penalty threat from antitrust authorities.
\((\psi_0, \psi_1, G)\) which are compatible with the observed distribution of \((X, Z, D, Y)\) and satisfy the constraints imposed by the extended Roy model presented in the preceding section. In the sequel, we will rely on the following condition.

**Assumption 3.1 (Exogeneity) \((X, Z) \perp (\varepsilon_0, \varepsilon_1)\)**

Assumption 3.1 states that the covariates \(X\) and \(Z\) affecting respectively the sector-specific potential earnings and the related non-pecuniary components are independent of the sector-specific unobserved productivity terms \((\varepsilon_0, \varepsilon_1)\). With respect to the covariates \(X\), it is noteworthy that the exogeneity assumption is very common in the literature addressing the identifiability of the Roy model (see e.g. Heckman and Honoré, 1990).\(^{19}\)

### 3.1 Identification of \((\psi_0, \psi_1)\)

#### 3.1.1 Identification without exclusion restrictions

In this subsection, we show that \(\psi_0\) and \(\psi_1\) can be identified without any exclusion restriction, by exploiting the structure of the generalized Roy model and under technical restrictions on the distribution of \((\varepsilon_0, \varepsilon_1)\). The idea behind is similar to the one of Heckman & Honore (1989) on the identification of competing risks models. There is indeed a strong analogy between competing risks, where only the identified minimum of several durations is observed, and standard Roy models, as it suffices to let \(\bar{Y}_i = 1/Y_i\) in the latter case to retrieve the framework of the former. Nevertheless, in order to obtain their identification results, Heckman & Honore (1989) rely on technical assumptions which are more restrictive than ours.\(^{20}\)

**Assumption 3.2 (Normalization) There exists \(x^*\) such that \(\psi_0(x^*) = \psi_1(x^*) = 0.\)**

\(^{19}\)Note that the exogeneity assumption may hold even if \(X\) (resp. \(Z\)) is endogenous, in which case \(\psi_0(.)\) and \(\psi_1(.)\) (resp. \(G(.)\)) are not structural.

\(^{20}\)More precisely, theorem 1 of Heckman & Honore (1989) crucially relies on assumption (i) which, in our context, amounts to assume that for all sequences \((u_{n,0})_n\) and \((u_{n,1})_n\) tending to infinity,

\[
\lim_{n \to \infty} e^{2u_{n,0}} \frac{\partial}{\partial u_{n,0}} P(\varepsilon_0 \leq u_{n,0}, \varepsilon_1 \leq u_{n,1}) = \lim_{n \to \infty} e^{2u_{n,1}} \frac{\partial}{\partial u_{n,1}} P(\varepsilon_0 \leq u_{n,0}, \varepsilon_1 \leq u_{n,1}) = l > 0 \quad (3.1)
\]

It can be shown that this condition implies in particular \(E(\exp(\varepsilon_0)) = E(\exp(\varepsilon_1)) = \infty\), so that \(\varepsilon_0\) and \(\varepsilon_1\) must have fat tails. Moreover, if \((\varepsilon_0, \varepsilon_1)\) are independent, (3.1) is equivalent to \(\lim_{x \to -\infty} e^{2x} f_{\varepsilon_0}(x) = \lim_{x \to -\infty} e^{2x} f_{\varepsilon_1}(x) = l\), so that only very peculiar densities will satisfy this requirement. Theorem 3.1, in contrast, will hold with \((\varepsilon_0, \varepsilon_1)\) independent under the only condition that \(E(\exp(\beta \varepsilon_0)) < +\infty\) and \(E(\exp(\beta \varepsilon_1)) < +\infty\) for a given \(\beta > 0.\)
Assumption 3.3 (Moderate dependence) The supports of $\varepsilon_0$ and $\varepsilon_1$ have infinite upper bounds, there exists $\beta > 0$ such that $E(\exp(\beta \varepsilon_0)) < +\infty$ and $E(\exp(\beta \varepsilon_1)) < +\infty$ and, for all $(\alpha_0, \alpha_1) \in \mathbb{R}^2$ and $i \in \{0, 1\}$,

$$\lim_{u \to \infty} P(\varepsilon_i \leq \alpha_i + u | \varepsilon_{1-i} = u) = 1.$$ 

Assumption 3.2 is an innocuous normalization which stems from the fact that adding a constant to $\psi_i$ and subtracting it to $\varepsilon_i$ does not modify the observables of the model. We then consider a set of restrictions on the distribution of $(\varepsilon_0, \varepsilon_1)$ which are detailed in Assumption 3.3. The first two restrictions are mainly technical. The second is a mild restriction on the tails of the distributions of $\varepsilon_0$ and $\varepsilon_1$, which is equivalent to $E[Y_i^\beta | X, Z] < \infty$ for at least one $\beta > 0$. Hence, in the case of sectorial choice, it will hold even if wages have very fat tails, Pareto like for instance. The last condition is more restrictive. It implies that there is no selection “at infinity”, that is to say

$$\lim_{y \to \infty} P(D = i | Y_i = y, X = x, Z = z) = 1$$

for all $(x, z)$ and $i \in \{0, 1\}$. The assumption holds when $\varepsilon_0$ and $\varepsilon_1$ are independent, and more generally as soon as the dependence between $\varepsilon_0$ and $\varepsilon_1$ remains moderately positive. When $(\varepsilon_0, \varepsilon_1)$ is normal bivariate for instance, one can show that it is satisfied when $\text{cov}(\varepsilon_0, \varepsilon_1) < V(\varepsilon_0) \wedge V(\varepsilon_1)$. Hence, when $V(\varepsilon_0) = V(\varepsilon_1)$, the assumption is automatically satisfied except in the degenerated case where $\varepsilon_0 = \varepsilon_1$. The idea of obtaining identification at the upper bound of the distribution is similar to the identification of competing risks models at zero established by Heckman & Honore (1989) and Abbring & van den Berg (2003).\textsuperscript{21}

\textbf{Theorem 3.1} Suppose that assumptions 3.1-3.3 hold. Then $(\psi_0, \psi_1)$ are identified.

Theorem 3.1 shows that $\psi_0$ and $\psi_1$ are identified without any exclusion restriction on the regressors which affect $Y_0$, $Y_1$ or $G(.)$. The knowledge of $G(Z)$ or a large support assumption on the effects of $X$ are not needed either. The main condition is assumption 3.3, which secures identification of the model at infinity.

\textsuperscript{21}To ensure the absence of selection at zero, Heckman & Honore (1989) impose their assumption (i) discussed in footnote 21. Abbring & van den Berg (2003) rely directly on the structure of the mixed proportional hazard model they examine as well as on a first-order moment condition which is more restrictive than ours. Another related work on competing risks is the one of Lee (2006), who shows that in generalized accelerated time models, identification can be achieved “globally” and not only at zero. However, his strategy breaks down here because of the unknown shift parameter $G(Z)$.
3.1.2 Identification with exclusion restrictions

The previous identification result is appealing in that it does not require any exclusion restrictions. However, identification is achieved at infinity, so that we expect the estimators to have a slow rate of convergence, as for instance in Fermanian (2003). In this section, and in a similar spirit as in Heckman & Honore (1990), we show that such an argument at the limit is unnecessary when there are exclusion restrictions. In the sequel, we let $X = (X_0, X_1, X_c)$.

**Assumption 3.4** (Exclusion restriction 1) $\psi_0$ (resp. $\psi_1$) depends only on $(X_0, X_c)$ (resp. on $(X_1, X_c)$). Moreover, $(X_0, X_c)$ (resp. $(X_1, X_c)$) and $P(D = 1|X, Z)$ are measurably separated, that is, any function of $(X_0, X_c)$ (resp. of $(X_1, X_c)$) almost surely equal to a function of $P(D = 1|X, Z)$ is almost surely constant.

**Assumption 3.5** (Continuous unobserved heterogeneity) The distribution of $(\varepsilon_0, \varepsilon_1)$ admits a density with respect to the Lebesgue measure.

Basically, the measurable separation requirement of assumption 3.4 ensures that $\psi_0(X)$ (or $\psi_1(X)$) and $P(D = 1|X, Z)$ can vary in a sufficiently independent way.\footnote{See Florens et al. (2008) for a discussion on this assumption.} Assumption 3.4 can cover two rather different situations. The first one is when $X_0 = X_1 = \emptyset$ but we observe some variables which affect the nonpecuniary components but not potential wages. Thus, this situation is close to the commonality condition of Bayer et al. (2008). The other one is when we observe some variables $X_0$ and $X_1$ which affect only one sector. In this latter case, no exclusion restriction on $Z$ is required. Assumption 3.5 is a technical condition which is usual in competing risks or Roy models (see among others Heckman & Honore, 1990, in the context of Roy model and Lee, 2006 in the context of competing risks models.). It implies that the distribution of $Y$ has no point mass.

Given the preceding exclusion restrictions and the exogeneity assumption, it is possible to identify $\psi_0(.)$ and $\psi_1(.)$ up to location parameters. Then full identification is achieved by the normalization of assumption 3.2.

**Theorem 3.2** Suppose that assumptions 3.1, 3.2, 3.4 and 3.5 hold. Then $\psi_0$ and $\psi_1$ are identified.
3.2 Identification of the nonpecuniary component

3.2.1 General analysis

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

**Assumption 3.6** *(Exclusion restriction 2)* For almost all $z$ in the support of $Z$, the distribution of $T$ conditional on $Z = z$ is not degenerated.

**Assumption 3.7** *(Large support on unobserved heterogeneity)* $(\epsilon_0, \epsilon_1)$ admits a continuous density $f_{\epsilon_0, \epsilon_1}$ and its support is $\mathbb{R}^2$.

Assumption 3.6 will allow us to make $X$ vary while holding the nonpecuniary component $G(Z)$ fixed. Hence, for assumption 3.6 to be verified, we need a variable which determines sector-specific potential earnings but does not directly affect the nonpecuniary component. This is the opposite of the kind of exclusion restrictions which are most often used to identify labor supply and more generally switching regressions models. For instance, in the context of college attendance decision considered in Section 2, local economic conditions (e.g. average local unemployment rate and average local wage) are natural candidates for these exclusion restrictions. Assumption 3.7 ensures that unobserved heterogeneity can be arbitrarily large. This assumption is also made by, e.g., Heckman & Honore (1990), and can be weaken without affecting much of our results, at the cost of complicating our arguments (see the discussion below theorem 3.3).

In the sequel, we omit the dependence in $Z$ for the ease of notation. Thus all the results must be understood for a given $z$ in the support of $Z$. For all real measurable function $h$ and $(t, \lambda) \in \mathbb{R}^2$, denoting by $F_{\epsilon_0|x} (\cdot)$ (resp. $F_{\epsilon_1|x}$) the cumulative distribution function of $\epsilon_0$ conditional on $\epsilon_1$ (resp. $\epsilon_1$ conditional on $\epsilon_0$), we let

$$\varphi_h(t, \lambda) = E \left[ F_{\epsilon_0|x} (\epsilon_1 - G - t|\epsilon_1) h(\epsilon_1 - \lambda) + F_{\epsilon_1|x} (\epsilon_0 + G + t|\epsilon_0) h(\epsilon_0) \right] .$$

(3.2)

Let $S$ denote the support of $T$. By the law of iterated expectation and assumption 3.1, $\varphi_h(t, \lambda)$ also writes, for all $t \in S$,

$$\varphi_h(t, \lambda) = E \left[ Dh(\epsilon - \lambda) + (1 - D)h(\epsilon)|T = t \right] .$$

\textsuperscript{23}Among others, Carneiro et al. (2003) exploit such exclusion restrictions in order to identify an extension of the Willis and Rosen's(1979) model of demand for college attendance.
In other terms, $\varphi_h(., \lambda)$ is identified on $S$. Now, by (3.2) and a change of variable, we get

$$
\frac{\partial \varphi_h}{\partial t}(t, \lambda) = \int_{\varepsilon_0 \varepsilon_1} f_{\varepsilon_0 \varepsilon_1}(v, v + G + t) [h(v) - h(v + G - \lambda + t)] dv
$$

(3.3)

At this stage, we use the following result, which is proved in the appendix.

**Lemma 3.1** Let $w(.)$ denote a positive function on the real line. For all $h \in I$, the set of nondecreasing and nonconstant functions, and all $\alpha \in \mathbb{R}$, we have

$$
\text{sgn} \left[ \int w(v)(h(v) - h(v - \alpha)) dv \right] = \text{sgn}(\alpha).
$$

The main point of lemma 3.1 is that the integral is strictly positive (resp. strictly negative) when $\alpha > 0$. This lemma, together with assumption 3.7, ensures that for any nondecreasing and nonconstant function $h$, $\varphi_h(., \lambda)$ is strictly increasing on $]-\infty, \lambda - G]$ and strictly decreasing on $[\lambda - G, +\infty[$. This property can be best illustrated by taking $h(t) = 1\{t \geq u\}$. When $t < \lambda - G$, $\varphi_h(t, \lambda)$ is equal to the area of $A_1 \cup A_2$ on the left of figure 3.2.1, weighted by the probability measure of $(\varepsilon_0, \varepsilon_1)$. When $t = \lambda - G$, $\varphi_h(t, \lambda)$ corresponds to the area of $B_1 \cup B_2$, whereas it is equal to the area of $C_1 \cup C_2$ when $t > \lambda - G$. Thus, $\varphi_h$ is strictly increasing on $]-\infty, \lambda - G]$ and strictly decreasing afterwards.

![Figure 1](image)

**Figure 1**: Properties of $\varphi_h(., \lambda)$ when $h(t) = 1\{t \geq u\}$.

By equation (3.3) and lemma 3.1, $\partial \varphi_h / \partial t(t, \lambda) = 0$ if and only if $\lambda = G + t$. Because $\partial \varphi_h / \partial t(., \lambda)$ is identified on $S$ when $T$ is continuous, $G$ is identified in this case. The situation is more involved when $T$ is discrete, as $\partial \varphi_h / \partial t(., \lambda)$ is not identified anymore. On the other hand, the properties of $\varphi_h(., \lambda)$ provide information on $G$. Indeed, given
\((t_i, t_j) \in S^2\) such that \(\lambda - G \leq t_i < t_j\), then for all \(h \in \mathcal{I}\), \(\varphi_h(t_i, \lambda) > \varphi_h(t_j, \lambda)\). Hence, if we observe \((t_i, t_j) \in S^2\) such that \(t_i < t_j\) and for a given \(h \in \mathcal{I}\), \(\varphi_h(t_i, \lambda) \leq \varphi_h(t_j, \lambda)\), then \(t_i < \lambda - G\), i.e., \(G < \lambda - t_i\). By optimizing over \((t_i, \lambda, h)\), we obtain the following upper bound on \(G\):

\[
G \leq \overline{G} = \inf_{(\lambda, t) \in \overline{\mathcal{D}}} \lambda - t
\]

where

\[
\overline{\mathcal{D}} = \{(\lambda, t) \in \mathbb{R} \times S / \exists (t', h) \in S \cap t, +\infty \times \mathcal{I} / \varphi_h(t, \lambda) \leq \varphi_h(t', \lambda)\}.
\]

A lower bound can be obtained similarly. A practical limitation of this expression is that as such, the set \(\mathcal{D}\) is difficult to impossible to implement, as it requires to compute \(\varphi_h\) for all \(h \in \mathcal{I}\). Actually, we show in the appendix that it suffices to restrict ourselves to the much smaller set \(\mathcal{I}_0 = \{t \mapsto \mathbb{1}\{t \geq u\}, u \in \mathbb{R}\}\) of indicator functions, yielding the following result.

**Theorem 3.3** Suppose that \((\psi_0, \psi_1)\) are identified and assumptions 3.1, 3.6 and 3.7 hold. Then:

\[
\sup_{(\lambda, t) \in \mathcal{D}} \lambda - t = \underline{G} \leq G \leq \overline{G} = \inf_{(\lambda, t) \in \overline{\mathcal{D}}} \lambda - t
\]

where

\[
\mathcal{D} = \{(\lambda, t) \in \mathbb{R} \times S / \exists (t', h) \in S \cap t, -\infty \times \mathcal{I}_0 / \varphi_h(t, \lambda) \leq \varphi_h(t', \lambda)\}, \quad \overline{\mathcal{D}} = \{(\lambda, t) \in \mathbb{R} \times S / \exists (t', h) \in S \cap t, +\infty \times \mathcal{I}_0 / \varphi_h(t, \lambda) \leq \varphi_h(t', \lambda)\}.
\]

Moreover, when \(T\) is continuous, \(G\) is identified.

Theorem 3.3 relies on the large support condition of assumption 3.7, which may seem restrictive. Instead, suppose the much weaker condition that there exists \((v, t) \in \mathbb{R} \times S\) such that \(f_{\epsilon_0, \epsilon_1}(v, v + G + t) > 0\). Then equation (3.3) shows that for all strictly increasing function \(h\), \(\partial \varphi_h / \partial t(t, \lambda) = 0\) if and only if \(\lambda = G + t\). Thus, \(G\) is still point identified when \(T\) is continuous. Similarly, lemma 3.1 remains valid for all strictly increasing function \(h\). Thus, the bound 3.4 still holds if we replace \(\mathcal{I}\) in \(\mathcal{D}\) by the set of strictly increasing function. However, we do not know whether this set can be restricted in this case to a smaller one.

### 3.2.2 An explicit method

The bounds given by (3.5) should be fairly tight in practice, as they exploit the information on \(\varphi_h\) for all \(h \in \mathcal{I}_0\). On the other hand, because of their complexity, it seems difficult to derive estimators of these bounds. Similarly, theorem 3.3 does not provide an explicit
In other terms, expression of $G$ which could be useful for inference when $T$ is continuous. We show here that using $h(t) = t$ enables to overcome these two issues. Let

$$
\begin{align*}
g_0(t) &= E \left[ F_{\varepsilon|t} (\varepsilon_1 - G - t|\varepsilon_1) \varepsilon_1 + F_{\varepsilon_0|t} (\varepsilon_0 + G + t|\varepsilon_0) \varepsilon_0 \right] \\
g_0(t) &= E \left[ F_{\varepsilon_0|t} (\varepsilon_1 - G - t|\varepsilon_1) \right]
\end{align*}
$$

When $t \in S$, these two functions simply reduce to $E(\varepsilon|T = t)$ and $E(D|T = t)$ respectively. Because $\partial \varphi_h / \partial t(t, G + t)$ is continuous, we have

$$
g_0(t) - (G + t)g_0'(t) = 0
$$

Integrating over $t$ and using an integration by part, we obtain, for any $t' \in S$,

$$
g_0(t) - tq_0(t) + \int_{t'}^t q_0(u)du = K + Gq_0(t)
$$

(3.6)

where $K = g_0(t') - t'q_0(t') - Gq_0(t')$. Because this equation holds for all $t \in S$, we have

$$
\varepsilon - DT + \int_{t'}^T q_0(u)du = K + DG + \eta, \quad E(\eta|T) = 0
$$

(3.7)

First, consider the case where the distribution of $T$ is continuous and $S$ is a compact interval. Then $q_0(u)$ is identified by $E(D|T = u)$ for all $u \in (t', T)$. Thus, it is possible to estimate nonparametrically the left hand side of (3.7). Once this has been done, estimation of $G$ is straightforward since it reduces to the estimation of a linear instrumental equation with one covariate.

When $T$ is discrete with finite support, the left hand side of (3.7) is not point identified anymore. However, equation (3.6) is still useful for inference. Indeed, it is easy to see that $q_0(.)$ is a decreasing function. Hence letting $t_1 < ... < t_M$ denote the points in $S$, we get, for all $i < j$ in $\{1, ..., M\}^2$,

$$
\sum_{k=i+1}^j (t_k - t_{k-1})q_0(t_k) \leq \int_{t_i}^{t_j} q_0(u)du \leq \sum_{k=i+1}^j (t_k - t_{k-1})q_0(t_{k-1}).
$$

Thus, for all $i < j$ in $\{1, ..., M\}^2$,

$$
\begin{align*}
&\quad \frac{g_0(t_j) - g_0(t_i) - t_jq_0(t_j) + t_iq_0(t_i) + \sum_{k=i+1}^j (t_k - t_{k-1})q_0(t_{k-1})}{q_0(t_j) - q_0(t_i)} \\
\quad \leq \quad G \leq \quad &\quad \frac{g_0(t_j) - g_0(t_i) - t_jq_0(t_j) + t_iq_0(t_i) + \sum_{k=i+1}^j (t_k - t_{k-1})q_0(t_{k-1})}{q_0(t_j) - q_0(t_i)}
\end{align*}
$$

In other terms,

$$
\begin{align*}
\max_{i<j} \left\{ \frac{g_0(t_j) - g_0(t_i) - t_jq_0(t_j) + t_iq_0(t_i) + \sum_{k=i+1}^j (t_k - t_{k-1})q_0(t_{k-1})}{q_0(t_j) - q_0(t_i)} \right\} \\
\leq \quad G \leq \quad \min_{i<j} \left\{ \frac{g_0(t) - g_0(t_i) - t_jg_0(t_j) + t_iq_0(t_i) + \sum_{k=i+1}^j (t_k - t_{k-1})q_0(t_{k-1})}{q_0(t_j) - q_0(t_i)} \right\}
\end{align*}
$$

16
This expression shows that when $M = 2$, the length of the interval will be exactly equal to $t_2 - t_1$. Hence, contrarily to many other examples in econometrics, large variations in the data are not desirable for identifying $G$.\footnote{On the other hand, such large variations may improve the accuracy of estimators.}

In general, the bounds obtained with this method will be strictly larger than the one defined by (3.5). Using lemma 8.1 (given in the Appendix) in the case where $M = 2$, we can build a sequence of distributions whose bounds defined by (3.5) shrink towards zero, whereas the bounds considered here will have a constant width. However, when $(\varepsilon_0, \varepsilon_1)$ is normal with mean $\mu$ and variance $\Sigma$ such that $\Sigma_{11} = \Sigma_{22}$, for instance, a simulation shows that the bounds coincide.\footnote{To compute the bounds defined by (3.5), we used numerical approximations of $\Delta P_{t,t,G,\lambda}(.)$ as defined in lemma 8.1.} Indeed, the length of the interval corresponding to (3.5) is exactly equal to $t_2 - t_1$ whatever the setting is (see Table 1 below). The basic setting corresponds to $\mu_0 = \mu_1 = 0$, $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = 0$, $G = 1$ and $S = \{t_1 = 0, t_2 = 1\}$. In this example, bounds appear to be sensitive to the correlation between $\varepsilon_0$ and $\varepsilon_1$. Increasing this correlation improves the upper bound, at the detriment of the lower bound.

<table>
<thead>
<tr>
<th>Basing setting, except that...</th>
<th>$G$</th>
<th>$\bar{G}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(basic setting)</td>
<td>0.44</td>
<td>1.44</td>
</tr>
<tr>
<td>$t_2 = 0.5$</td>
<td>0.74</td>
<td>1.24</td>
</tr>
<tr>
<td>$t_2 = 0.05$</td>
<td>0.97</td>
<td>1.02</td>
</tr>
<tr>
<td>$\mu_0 = 1$</td>
<td>0.40</td>
<td>1.40</td>
</tr>
<tr>
<td>$\Sigma_{12} = 0.5$</td>
<td>0.38</td>
<td>1.38</td>
</tr>
<tr>
<td>$\Sigma_{12} = 0.9$</td>
<td>0.16</td>
<td>1.16</td>
</tr>
<tr>
<td>$\Sigma_{12} = 0.99$</td>
<td>0.02</td>
<td>1.02</td>
</tr>
<tr>
<td>$G = 2$</td>
<td>1.40</td>
<td>2.40</td>
</tr>
</tbody>
</table>

Table 1: Bounds on $G$ given by (3.5) when $(\varepsilon_0, \varepsilon_1)$ is normal

### 4 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 generalized Roy models...
with index structures of the form:\textsuperscript{26}

\[
\ln(Y_0) = X'\beta_0 + \varepsilon_0 \\
\ln(Y_1) = X'\beta_1 + \varepsilon_1 \\
D = 1\{\ln(Y_1) > \ln(Y_0) + \delta_0 + X'\gamma_0\}.
\]

In this setting, the nonpecuniary component is of the form \(\delta_0 + X'\gamma_0\). Let \(\gamma_{0k}\) (resp. \(\beta_{0k}\), \(\beta_{1k}\)) denote the \(k\)-th component of \(\gamma_0\) (resp. \(\beta_0\), \(\beta_1\)). We impose the following assumptions, which correspond to the exclusion restrictions of assumptions 3.4 and 3.6, as well as the continuity of \(T\).

**Assumption 4.1** (Exclusion restrictions) We have \(\beta_{01} = \beta_{12} = 0\), \(\gamma_{01} \neq \beta_{11}\) and \(\gamma_{02} \neq -\beta_{02}\). Moreover, there exists \(m\) such that \(\gamma_{0m} = 0\) and \(\beta_{0m} \neq \beta_{1m}\).

**Assumption 4.2** (Regularity of \(X\)) The support of \(X\) is bounded. Moreover, for all \(x_{-m}\) in the support of \(X_{-m} = (X_1, \ldots, X_{m-1}, X_{m+1}, \ldots)\), the distribution of \(X_m\) conditional on \(X_{-m} = x_{-m}\) admits a continuously differentiable positive density on its support, which is a compact interval. Lastly, for all \(j\), \(u \mapsto E(X_j|T = u)\) is continuously differentiable.

Assumption 4.2 states that the conditional density of the continuous instrument is positive on its support, which is a compact interval. Actually, estimation would be simpler if the distribution of \((X'\beta_0, X'\beta_1, X'\zeta_0)\) has had a large support, since then the nuisance parameter, i.e. the joint distribution of \((\varepsilon_0, \varepsilon_1)\), would be identified. In that case one could rely on sieve estimation (see e.g. Chen (2007)).

We propose a three-stages estimation procedure of the preceding model based on a sample of \((Y = D Y_1 + (1 - D) Y_0, X, D)\).

**Assumption 4.3** (i.i.d. sample) We observe a sample \((Y_i, X_i, D_i)_{1 \leq i \leq n}\) of i.i.d. copies of \((Y, X, D)\).

Let us assume, without loss of generality, that \(\beta_{1m} - \beta_{0m}\) is strictly positive. We define \(\zeta_0 = (\beta_1 - \beta_0 - \gamma_0)/(\beta_{1m} - \beta_{0m})\) and \(\nu = (\varepsilon_1 - \varepsilon_0 - \delta_0)/(\beta_{1m} - \beta_{0m})\). The first and second stages of our procedure rely on the fact that we can rewrite the model as

\[
\begin{align*}
D &= 1\{X'\zeta_0 + \nu > 0\} \\
\ln(Y_k) &= X'\beta_k + \varepsilon_k \quad (k \in \{0, 1\})
\end{align*}
\]

\textsuperscript{26}We suppose that the constant is not included in \(X\), so that \(\varepsilon_0\) and \(\varepsilon_1\) do not necessarily have mean zero.
where $Y_k$ is observed when $D = k$, $(\nu, \varepsilon_k)$ is independent of $X$ by assumption 3.1 and, by assumption 4.1, at least one of the component in $X'\zeta_0$ is not included in $X'\beta_k$. Hence, Equation (4.1) corresponds to Newey (2008)'s selection model and we follow his approach here. First, we estimate $\zeta_0$ by any single index estimator proposed in the literature (see e.g. Klein and Spady, 1993, Horowitz & Hardle, 1996 or Ichimura, 1993). We impose the following condition on this estimator $\hat{\zeta}$. This condition is satisfied by many semiparametric estimators of binary choice models, such as the one of Klein & Spady (1993).

**Assumption 4.4 (Regularity of the first stage estimator)** There exists $(\psi_i)_{1 \leq i \leq n}$, i.i.d. random variables such that $E(\psi_1) = 0$, $E(\psi_1 \psi_1')$ exists and is non singular and

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

Secondly, we estimate $\beta_0$ and $\beta_1$ by series estimator. We impose the following assumption on the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. Note that it is possible to prove this condition under more primitive assumptions (see Newey, 2008, p.10).

**Assumption 4.5 (Regularity of the second stage estimators)** Let $k \in \{0, 1\}$, there exists $(\psi_{ki})_{1 \leq i \leq n}$, i.i.d. random variables such that $E(\psi_{k1}) = 0$, $E(\psi_{k1} \psi_{k1}')$ exists and is non singular and

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

Once $(\zeta_0, \beta_0, \beta_1)$ have been estimated, we can easily recover $\gamma_0$. Since $\gamma_0 = \beta_1 - \beta_0 - \zeta_0(\beta_{1m} - \beta_{0m})$, we let

$$
\hat{\gamma} = \hat{\beta}_1 - \hat{\beta}_0 - \hat{\zeta}(\beta_{1m} - \beta_{0m}).
$$

It is easy to see that assumption 4.4 and 4.5 imply the root-n convergence and asymptotic normality of $\hat{\gamma}$. Thus, the main difficulty lies in the estimation of $\delta_0$, which we now consider.

Let $\xi_0 = \beta_0 - \beta_1 + \gamma_0$ and define, as before, $T_i = X'_i\xi_0$. The third stage of our procedure is based on (3.7), which writes as

$$
\varepsilon_i - D_iT_i + \int_{t_0}^{T_i} q_0(u)du = K + D_i\delta_0 + \eta_i, \quad E(\eta_i|T_i) = 0.
$$

Let $\theta_0 = (K, \delta_0)'$, $V_i = \varepsilon_i - D_iT_i + \int_{t_0}^{T_i} q_0(u)du$, $W_i = (1, D_i)'$ and $S_i = (1, h(T_i))'$ for a given function $h$. Then $V_i = W_i'\theta_0 + \eta_i$ and $\theta_0 = E(S_iW_i')^{-1}E(S_iV_i)$. We estimate $\theta_0$ by

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

19
where

\[ \hat{V}_i = \hat{\varepsilon}_i - D_i \hat{T}_i + \int_{t_0}^{\hat{T}_i} \hat{q}(u, \xi) \, du \]

with \( \hat{\varepsilon}_i = \ln Y_i - X_i' (D_i \hat{\beta}_1 + (1 - D_i) \hat{\beta}_0) \), \( \xi = \hat{\beta}_0 - \hat{\beta}_1 + \hat{\gamma} \), \( \hat{T}_i = X_i' \hat{\xi} \) and

\[ \hat{q}(u, \xi) = \frac{\sum_{i=1}^{n} D_i K \left( \frac{u - X'_i \xi}{h_n} \right)}{\sum_{i=1}^{n} K \left( \frac{u - X'_i \xi}{h_n} \right)}. \]

The result on the third step estimator \( \hat{\theta} \) relies on the following conditions.

**Assumption 4.6** (Regularity on \((\varepsilon_0, \varepsilon_1)\)) The distribution of \( \varepsilon_0 - \varepsilon_1 \) admits a continuous density with respect to the Lebesgue distribution. Moreover, \( E(|\varepsilon_0|) < +\infty \) and \( E(|\varepsilon_1|) < +\infty \).

**Assumption 4.7** (Regular instruments) \( h(\cdot) \) is twice differentiable and \( h'(\cdot) \) and \( h''(\cdot) \) are bounded. Moreover, \( E(|h'^2(T_i)|) < +\infty \).

**Assumption 4.8** (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 \( \int v K(v) \, dv = 0 \). Moreover, \( K(\cdot) \) and \( K'(\cdot) \) are zero on the boundary of this compact set.

Assumption 4.6 is standard and fairly weak. Assumption 4.7 could probably be relaxed, but holds for instance when \( h(t) = t \), ie when taking \((1, T)\) as instruments. Lastly, assumption 4.8 is satisfied for instance by the quartic kernel \( K(v) = (15/16)(1 - v^2)^2 \mathbb{1}_{[-1,1]}(v) \).

**Theorem 4.1** Suppose that \( nh_n^6 \to \infty \), \( nh_n^8 \to 0 \) and assumptions 4.1-4.8 hold. Then

\[ \sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} \mathcal{N}(0, E(S_1 W_1' t)^{-1} V(\Omega_{11} + \Omega_{21} + \eta S_1) E(W_1 S_1')^{-1}) . \]

where \( \Omega_{11} \) is defined by equation (7.13) in the appendix and

\[ \Omega_{21} = S_T(T_i) \mathbb{1}\{T_i \geq t_0\} (D_i - q_0(T_i)) / f_0(T_i) \]

where \( S_T(\cdot) \) denotes the survival function of \( T_1 \).

This theorem guarantees that the final estimator of \( \delta_0 \) is root-n consistent and asymptotically normal. Its asymptotic variance depends on the three variables \( \Omega_{11} \), \( \Omega_{21} \) and \( \eta S_1 \). The first corresponds to the contribution of the first step estimators. The second arises because of the nonparametric estimation of \( q_0(\cdot) \). The third simply corresponds to the moment estimation of the linear model (3.7) in the last step.
5 Monte Carlo simulations

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

\[
\begin{align*}
\ln(Y_{0i}) &= \beta_{02} X_{2i} + \beta_{03} X_{3i} + \varepsilon_{0i} \\
\ln(Y_{1i}) &= \beta_{11} X_{1i} + \beta_{13} X_{3i} + \varepsilon_{1i} \\
D_i &= \mathbf{1}\{\ln(Y_{1i}) > \ln(Y_{0i}) + \delta_0 + \gamma_03 X_{3i}\}
\end{align*}
\]

\(X_{1i}\) and \(X_{2i}\) are uniformly distributed over \([0, 4]\), while \(X_{3i}\) is a discrete regressor which is drawn from a Bernoulli distribution with parameter \(p = 0.5\). The true values of the parameters are \(\beta_{02} = \beta_{03} = 1, \beta_{13} = 0.5, \beta_{11} = 2, \gamma_03 = -0.8\) and \(\delta_0 = 0.8\). We finally let \((\varepsilon_{0i}, \varepsilon_{1i})'\) be joint normal, with mean \(\mu = (0, 0)'\) and variance \(\Sigma\) such that \(\Sigma_{11} = \Sigma_{22} = 1\) and \(\Sigma_{12} = \Sigma_{21} = 0.5\).

We implement the three-stages estimation procedure detailed in Section 4. More precisely, in the first stage, we estimate \(\zeta_0 = (\beta_1 - \beta_0 - \gamma_0)/\beta_{11}\) (since \(m = 1\) and \(\beta_{0m} = 0\) here) by implementing Klein & Spady’s (1993) semiparametric estimation method with an adaptive gaussian kernel with local smoothing. In the second stage, we implement Newey’s (2008) method in order to estimate separately \(\beta_0, \beta_1\) and \(\gamma_0\). The series estimator of the selection correction term was computed using Legendre polynomials at order 6. In the third stage, we finally implement our proposed estimator for \(\delta_0\) with the quartic kernel suggested in Section 4, namely \(K(v) = (15/16)(1 - v^2)^2 \mathbf{1}_{[-1,1]}(v)\).

The performance of the estimators (with sample size \(N = 1,000\) and \(1,000\) Monte Carlo replications) are summarized in Table 2, which report for each parameter the mean estimate, the standard deviation and the root-mean-squared error. The results indicate that the semiparametric estimation procedure proposed in Section 4 performs relatively well in this context. In particular, although the last-stage estimator of the non-pecuniary constant component \(\hat{\delta}_0\) is as expected less precise than the estimators \(\hat{\beta}_0, \hat{\beta}_1\) and \(\hat{\gamma}_0\) (with a RMSE of 0.298), its finite-sample performance still remain reasonable.

\(^{27}\)Note that it is possible to apply our estimation strategy to a more general framework, either with \(\gamma_{01} \neq 0\) or \(\gamma_{02} \neq 0\). Nevertheless, the data generating process we consider here has the advantage of exploiting the same kind of exclusion restrictions as the ones used in our application.
### Table 2: Monte Carlo simulations

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{02}$</td>
<td>1.004</td>
<td>0.110</td>
<td>0.110</td>
</tr>
<tr>
<td>$\hat{\beta}_{03}$</td>
<td>0.991</td>
<td>0.057</td>
<td>0.058</td>
</tr>
<tr>
<td>$\hat{\beta}_{11}$</td>
<td>1.998</td>
<td>0.078</td>
<td>0.078</td>
</tr>
<tr>
<td>$\hat{\beta}_{13}$</td>
<td>0.515</td>
<td>0.078</td>
<td>0.079</td>
</tr>
<tr>
<td>$\hat{\gamma}_{03}$</td>
<td>-0.791</td>
<td>0.177</td>
<td>0.177</td>
</tr>
<tr>
<td>$\hat{\delta}_{0}$</td>
<td>0.843</td>
<td>0.295</td>
<td>0.298</td>
</tr>
</tbody>
</table>

6 Application to the decision to attend higher education

6.1 Data

We use French data from the *Generation 1992* and *Generation 1998* surveys in order to estimate a model of schooling choice, applied to the decision to attend higher education. The *Generation 1992* survey consists of a large sample of 26,359 individuals who left the French educational system in 1992 and were interviewed five years later, in 1997. This database has the main advantage to contain information on both educational and labor market histories (over the first five years following the exit from the educational system). Furthermore, the survey provides a set of individual covariates which are used as controls in our estimation procedure such as gender, place of birth, nationality, parents’ profession, and place of residence. Most of the individual covariates observed in the *Generation 1992* survey are also provided by the *Generation 1998* survey, which consists of a sample of 22,021 individuals who left the French educational system six years later, in 1998, and were interviewed in 2003. In this paper, we exploit the pooled dataset which contains informations on a total of 48,380 individuals entering the labor market either in 1992 or in 1998.

Our subsample of interest is constituted of respondents from both of these surveys having at least passed the national high school final examination. Dropping individuals who only worked as temporary workers or houseworkers during the observation length, for whom wages are not observed in the data, finally leaves us with a sample of 24,676 individuals.

---

28 These data have been previously used in the educational context by Beffy et al. (2009) as well as Brodaty et al. (2009).

29 Note that labor market participation rates are fairly high for the subsample we consider in the paper. Namely, for individuals leaving the schooling system in 1992, they are equal to 99.7% for males and 95.9%
We report below some descriptive statistics for the subsample of interest, according to the year of entry into the labor market. Respectively 77.5% (for Generation 1992) and 80.8% (for Generation 1998) of our sample attended higher education after graduating from high school.

6.2 The model

We rely on the generalized Roy framework which is presented in subsection 2.2.1. Individuals are supposed to make their schooling decision (either high school, $c^* = 0$, or higher education attendance, $c^* = 1$) by comparing the expected streams of earnings related to each alternative. For a given period $t$, the earnings variable 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.

Denoting by $y_{c,t}$ the log-earnings for each (either employment or unemployment) spell $t$ and each alternative $c$, the earnings dynamics we posit is similar as the one considered in subsection 2.2.1, that is, for each employment spell $t$:

$$y_{c,t} = \rho_c L_{t-1} + y_{c,t-1} + \varepsilon_{c,t}$$

Where $\rho_c$ denotes the alternative $c$-specific return to experience, $L_{t-1}$ is the length of the $(t-1)^{th}$ spell (in years), and $\varepsilon_{c,t}$ is a degree $c$-specific unobserved individual productivity term which is assumed to be independently and identically distributed over time, with mean zero. We estimate $\rho_0$ and $\rho_1$ by pooled OLS respectively on the subsample of high school and higher education graduates. We further assume that initial earnings $y_{c,0}$ are given by:

$$y_{c,0} = u_c + \varepsilon_{c,0}$$

Where $u_c = \psi_c(X) + \eta_c$ is an individual term depending on the schooling choice $c$. $\psi_c(.)$ are unknown functions of observed individual covariates $X$, while $\eta_c$ is a productivity random term which is supposed to be known by the individual at the time of her decision.

We further assume that the individuals decide on whether to attend higher education by comparing, for each schooling alternative, the expected stream of future labor market earnings. The indirect utility related to each schooling decision $c$ thus expresses as, denoting by $y_{c,\tau}$ the earnings received during the $\tau^{th}$ year of active life:

$$V_c = \sum_{\tau=0}^{T} \beta^\tau E(y_{c,\tau}|X, u_c) + G_c(Z)$$

for females, while for those leaving education in 1998, the participation rates are equal to 99.3% for males and 97.2% for females. Thus, we decide to keep both males and females in our final sample.
With:

\[ y_{c,T} = \tau \rho_c + u_c + \sum_{k=0}^{\tau} \varepsilon_{c,k} \]

Assuming a constant replacement rate \( e^{b} \), \( V_c \) can be rewritten as:

\[
V_c = \left( \sum_{t=0}^{T} t\beta^t \right) \rho_c + \left( \sum_{t=0}^{T} \beta^t \right) (\psi_c(X) + \eta_c) + G_c(Z)
\]

Finally, we will observe \( c^* = 1 \) (higher education attendance) in the data if and only if:

\[
\eta_1 > -\frac{G_1(Z)}{\sum_{t=0}^{T} \beta^t} + \frac{\sum_{t=0}^{T} t\beta^t}{\sum_{t=0}^{T} \beta^t} (\rho_0 - \rho_1) + (\psi_0 - \psi_1)(X) + \eta_0
\]

6.3 Results

To be completed

\(^{30}\)We set the replacement rate \( e^b = 0.7 \) as it is often done in the literature. Besides, for the sake of simplicity, we define \( b_0 = 0 \).
7 Appendix A: proofs

Theorem 3.1

We focus on \( \psi_1 \), as the proof is identical for \( \psi_0 \). Let \( y > 0 \), \((x, z)\) belong to the support of \((X, Z)\) and define \( q(y, x, z) = P(D = 1, Y \geq y | X = x, Z = z) \). We have

\[
q(y, x, z) = P(\varepsilon_0 \leq \psi_1(x) + \varepsilon_1 - G(z) - \psi_0(x), \varepsilon_1 \geq y - \psi_1(x)) = \int_{y-\psi_1(x)}^{\infty} P(\varepsilon_0 \leq \psi_1(x) + u - G(z) - \psi_0(x) | \varepsilon_1 = u) dP_{\varepsilon_1}(u) \tag{7.1}
\]

By assumption 3.3, as as \( u \to \infty \), we have

\[
P(\varepsilon_0 \leq \psi_1(x) + u - G(z) - \psi_0(x) | \varepsilon_1 = u) \to 1. \tag{7.2}
\]

Thus, using standard integral instruments, we get as \( y \to \infty \)

\[
q(y, x, z) \sim S_{\varepsilon_1}(y - \psi_1(x)).
\]

Similarly, letting \( z^* \) be such that \((x^*, z^*)\) belongs to the support of \((X, Z)\), with by assumption 3.2 \( \psi_1(x^*) = 0 \), we have

\[
q(y, x^*, z^*) \sim S_{\varepsilon_1}(y) \tag{7.3}
\]

In other words,

\[
q(y, x, z) \sim q(y - \psi_1(x), x^*, z^*). \]

Now, let us show that actually, as \( y \to \infty \),

\[
q(y, x, z) \sim q(y + u, x^*, z^*) \implies u = -\psi_1(x) \tag{7.4}
\]

Because the function \( q \) is identified, this implies that \( \psi_1(x) \) is identified. To prove (7.4), suppose that there exists \( u \neq -\psi_1(x) \) such that \( q(y, x, z) \sim q(y + u, x^*, z^*) \). Let \( v = u + \psi_1(x) \) if \( u > -\psi_1(x) \), \( v = -(u + \psi_1(x)) \) otherwise. Then, \( v > 0 \) and, as \( y \to \infty \),

\[
q(y + v, x^*, z^*) \sim q(y, x^*, z^*).
\]

This implies, by (7.3), that \( S_{\varepsilon_1}(v + y) \sim S_{\varepsilon_1}(y) \). Hence, \( S_{v\varepsilon_1}(wy) \sim S_{v\varepsilon_1}(y) \) where \( \beta \) is defined in assumption 3.3 and \( w = \exp(\beta v) > 1 \). Thus, letting \( \eta < 1 - 1/w \), there exists \( y_0 \) such that for all \( y > y_0 \), \( S_{v\varepsilon_1}(wy) > (1 - \eta)S_{v\varepsilon_1}(y) \). Now, since we have \( E(\exp(\beta \varepsilon_1)) = \int_0^{\infty} S_{v\varepsilon_1}(u) du \), it follows from assumption 3.3 that the integral of \( S_{v\varepsilon_1}(y) \) is finite. Thus,

\[
\int_{y_0}^{\infty} S_{v\varepsilon_1}(wy) dy > (1 - \eta) \int_{y_0}^{\infty} S_{v\varepsilon_1}(y) dy.
\]
This implies that
\[
\frac{1}{w} \int_{y_0}^{\infty} S_{e^{\beta y}}(y)dy > (1 - \eta) \int_{y_0}^{\infty} S_{e^{\beta y}}(y)dy > (1 - \eta) \int_{y_0}^{\infty} S_{e^{\beta y}}(y)dy.
\]
In other words, \(1/w > 1 - \eta\), which is a contradiction. Hence \(u\) is unique, and the result follows.

**Theorem 3.2**

First, we have, by assumptions 3.1 and 3.5,
\[
E(\varepsilon_1|D = 1, X = x, Z = z) = \frac{E(\varepsilon_1 D|x, Z = z)}{P(D = 1|x, Z = z)} = \frac{E(\varepsilon_1 \{\varepsilon_1 - \varepsilon_0 \geq \psi_0(x) - \psi_1(x) + G(z)\})}{P(D = 1|x, Z = z)}
\]
Now let us show that almost surely,
\[
\varepsilon_1 - \varepsilon_0 \geq \psi_0(x) - \psi_1(x) + G(z) \iff S_{\varepsilon_1 - \varepsilon_0}(\varepsilon_1 - \varepsilon_0) \leq P(D = 1|x, Z = z) \tag{7.6}
\]
where \(S_{\varepsilon_1 - \varepsilon_0}\) denotes the survival function of \(\varepsilon_1 - \varepsilon_0\). The first implication is obvious since \(S_{\varepsilon_1 - \varepsilon_0}\) is decreasing. Now suppose that \(S_{\varepsilon_1 - \varepsilon_0}(\varepsilon_1 - \varepsilon_0) \leq P(D = 1|x, Z = z)\).
Then \(\varepsilon_1 - \varepsilon_0 \geq \inf A_{x,z}\) where \(A_{x,z} = \{u/S_{\varepsilon_1 - \varepsilon_0}(u) = P(D = 1|x, Z = z)\}\). Now, for all interval \(I \subset A_{x,z}\), \(P(\varepsilon_1 - \varepsilon_0 \in I) = 0\) by definition of \(A_{x,z}\). Hence, because \(\psi_0(x) - \psi_1(x) + G(z) \in A_{x,z}\), almost surely,
\[
\varepsilon_1 - \varepsilon_0 \geq \inf A_{x,z} \Rightarrow \varepsilon_1 - \varepsilon_0 \geq \psi_0(x) - \psi_1(x) + G(z).
\]
Hence, (7.6) holds. Then, by (7.5),
\[
E(\varepsilon_1|D = 1, X = x, Z = z) = \frac{E(\varepsilon_1 \{S_{\varepsilon_1 - \varepsilon_0}(\varepsilon_1 - \varepsilon_0) \leq P(D = 1|x, Z = z)\})}{P(D = 1|x, Z = z)}
\]
In other terms, there exists a measurable function \(h\) such that \(E(\varepsilon_1|D = 1, X, Z) = h(P(D = 1|X, Z))\). Now, by assumption 3.4,
\[
E(\ln Y|D = 1, X) = \psi_1(X, X_e) + h(P(D = 1|X, Z)).
\]
Suppose that there exists \(\tilde{\psi}_1\) and \(\tilde{h}\) such that
\[
E(\ln Y|D = 1, X) = \tilde{\psi}_1(X, X_e) + \tilde{h}(P(D = 1|X, Z)).
\]
Then
\[
(\tilde{\psi}_1 - \psi_1)(X, X_e) + (\tilde{h} - h)(P(D = 1|X, Z)) = 0
\]
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 3.2. Thus, \(\tilde{\psi}_1\) is identified. \(\psi_0\) can be recovered by the same argument.
Proof of lemma 3.1

We prove the result only for \( \alpha > 0 \), as the reasoning is similar for \( \alpha < 0 \) and the equality is trivial for \( \alpha = 0 \). Because \( w(.) \) is a positive function, it suffices to show that there exists an interval \( I \) different from a singleton such that for all \( x \in I \),

\[
  h(x + \alpha) - h(x) > 0.
\]

Suppose the contrary. Because \( h \) is nonconstant, there exists \( x' > x \) such that \( h(x') > h(x) \). Let \( K \in \mathbb{N} \) be such that \( x' - x \leq K\alpha/2 \) and let \( 0 < \varepsilon < \alpha \). There exists \( x_0 \in [x - \alpha/2, x] \) such \( h(x_0 + \alpha) = h(x_0) \). Note that by monotonicity, this implies that \( h \) is constant on \([x_0, x_0 + \alpha]\). Similarly, there exists \( x_1 \in [x_0 + \alpha - \varepsilon, x_0 + \alpha] \) such \( h(x_1 + \alpha) = h(x_1) \). Moreover, by monotonicity and because \( x_1 + \alpha \geq x_0 + 2\alpha - \varepsilon \geq x_1 \), we get \( h(x_0 + 2\alpha - \varepsilon) = h(x_1) = h(x_0 + \alpha) \). Thus, \( h \) is constant on \([x_0, x_0 + 2\alpha - \varepsilon]\). A similar reasoning shows that for all \( n \in \mathbb{N}^* \), \( h \) is constant on \([x_0, x_0 + n\alpha - (n - 1)\varepsilon]\). Letting \( n = K \) and \( \varepsilon = \alpha/2 \) shows that \( h(x') = h(x_0) = h(x) \), a contradiction. The result follows.

Proof of theorem 3.3

We prove the result for the upper bound only, the reasoning being similar for the lower bound. Let \( h_u(t) = 1 \{ t \geq u \} \). By the discussion preceding the theorem, it suffices to show that for all \( \lambda \in \mathbb{R} \) and \( t < t' \in S \),

\[
  [\exists h \in \mathcal{I} / \varphi_h(t, \lambda) \leq \varphi_h(t', \lambda)] \implies [\exists u \in \mathbb{R} / \varphi_{h_u}(t, \lambda) \leq \varphi_{h_u}(t', \lambda)] \tag{7.7}
\]

To prove this result, suppose that for all \( u \in \mathbb{R} \), \( \varphi_{h_u}(t, \lambda) > \varphi_{h_u}(t', \lambda) \). First, note that \( t' > \lambda - G \), because otherwise \( \varphi_{h_u}(t, \lambda) < \varphi_{h_u}(t', \lambda) \) since \( \varphi_{h_u}(t, \lambda) \) is strictly increasing on \([-\infty, \lambda - G]\). Similarly, we can restrict our attention to \( t < \lambda - G \), since \( \varphi_{h}(t, \lambda) > \varphi_{h}(t', \lambda) \) is obvious for all \( h \in \mathcal{I} \) when \( t \geq \lambda - G \). Hence \( t < \lambda - G < t' \).

Now, let us define

\[
  \Delta P_{t,t',G,\lambda}(w) = P(\varepsilon_0 \geq w, \varepsilon_0 + t + G \leq \varepsilon_1 \leq w + \lambda) - P(\varepsilon_0 \leq w, \varepsilon_0 + t' + G \geq \varepsilon_1 \geq w + \lambda).
\]

By lemma 8.1, for all \( u \in \mathbb{R} \),

\[
  \Delta P_{t,t',G,\lambda}(u) = \int \Delta P_{t,t',G,\lambda}(w) dh_u(w) = \varphi_{h_u}(t', \lambda) - \varphi_{h_u}(t, \lambda) < 0.
\]

Hence, because \( \Delta P_{t,t',G,\lambda}(.) \) is continuous and functions in \( \mathcal{I} \) are nonconstant, we have for all \( h \in \mathcal{I} \)

\[
  \int \Delta P_{t,t',G,\lambda}(w) dh(w) < 0.
\]
Thus, by lemma 8.1 once more, \( \varphi_h(t, \lambda) > \varphi_h(t', \lambda) \). This shows (7.7), and the result follows.

**Proof of theorem 4.1**

Before proving the results, let us introduce some notations. Let \( U_i \) denote all the data corresponding to individual \( i \), let \( f_0(., \xi) \) denote the density of \( X' \xi \), \( q_0(u, \xi) = E(D|X'\xi = u) \), \( r_0(., \xi) = q_0(., \xi) \times f_0(., \xi) \) and \( \lambda_0 = (r_0, f_0, \xi_0, \beta_0, \beta_1) \). Let also

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

and \( \hat{\lambda} = (\hat{r}(., \xi), \hat{f}(., \xi), \hat{\xi}, \hat{\beta}_0, \hat{\beta}_1) \) where \( \hat{r}(., \xi) = \hat{q}(., \xi) \times \hat{f}(., \xi) \). Then we can write \( \hat{V}_i = V_i(\hat{\lambda}) \). Let us also define \( S_i(\xi) = (1, h(X'_i \xi))' \). Eventually, let \( U_i = X' i \xi \), \( g(U_i, \theta, \lambda) = S_i(\xi)(V_i(\lambda) - W'_i \theta) \) and \( g(U_i, \lambda) = g(U_i, \theta_0, \lambda) \) for any \( \lambda = (r, f, \xi, \beta_0, \beta_1) \). Then \( E[g(U_1, \lambda_0)] = 0 \) and

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

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. Note however that some differences arise because of the estimation of \( \xi \) in the nonparametric estimator of \( q_0 \). The proof of the theorem proceeds in three steps.

Step 1. We show that \( \lambda \mapsto \sum_{i=1}^{n} g(U_i, \lambda) \) can be linearized in a convenient way. For any \( \lambda = (r, f, \xi, \beta_0, \beta_1) \), let

\[
G(U_i, \lambda) = [V_i(\lambda_0) - W'_i \theta_0] \frac{\partial S_i}{\partial \xi}(\xi_0)' \xi + S_i(\xi_0) \left[ -X'_i(D_i \beta_1 + (1 - D_i) \beta_0) - D_i X'_i \xi 
\right.
\]

\[
+ q_0(T_i, \xi_0) X'_i \xi + \int_{T_i}^{T_i} \frac{\partial q_0}{\partial \xi}(u, \xi_0)' \xi + \frac{1}{f_0(u)} (r(u) - q_0(u)f(u)) du \right]
\]

Note that \( \frac{\partial q_0}{\partial \xi}(., \xi_0) \) exists under assumptions 4.2 and 4.6, by lemma 8.2. Let us also define \( \tilde{\lambda} = (\tilde{r}, \tilde{f}, \tilde{\xi}) \) where \( \tilde{r} = \tilde{r}(., \xi_0) \) and \( \tilde{f} = \tilde{f}(., \xi_0) \). We shall prove that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g(U_i, \tilde{\lambda}) - g(U_i, \lambda_0) - G(U_i, \tilde{\lambda} - \lambda_0) \right] = o_P(1).
\]

For that purpose, we use the decomposition

\[
g(U_i, \tilde{\lambda}) - g(U_i, \lambda_0) - G(U_i, \tilde{\lambda} - \lambda_0) = R_{1i} + R_{2i} + R_{3i} + R_{4i} + R_{5i}
\]

28
By assumption 4.4 and 4.5, the first term in the r.h.s. is a
\[ o \left| \frac{q}{\sqrt{n}} \right| - 1 \]
where \( q \) with \( q = \frac{\partial S_i}{\partial \xi} (\xi_0) (\hat{\xi} - \xi_0) \).

This implies that
\[ S_i(\xi) \]
Moreover, because \( S \) is defined by assumption 4.2 and the Cauchy-Schwartz inequality,
\[
\| S_i(\hat{\xi}) - S_i(\xi_0) \| = \left| h(X'_i \hat{\xi}) - h(X'_i \xi_0) - h'(X'_i \xi_0)X'_i(\hat{\xi} - \xi_0) \right|
\leq C_0 \left| X'_i(\hat{\xi} - \xi_0) \right|^2
\leq C_0 M \left\| \hat{\xi} - \xi_0 \right\|^2.

Moreover, because \( S \) is bounded, there exists \( C_1, C_2 > 0 \) such that for all \( i, |T_i| < C_1 \) and \( |T_i - t_0| < C_2 \). Moreover \( q_0(.) \) is bounded by 1. Thus,
\[
|V_i(\lambda_0)| \leq |\epsilon_i| + C_1 + C_2 \leq |\epsilon_0| + |\epsilon_{1i}| + C_1 + C_2
\]
Besides, \( |W'_i\theta_0| \leq \sqrt{2} \|\theta_0\| \). Thus,
\[
|V_i(\lambda_0) - W'_i\theta_0| \leq \sqrt{2} \|\theta_0\| \leq \sqrt{2} \|\theta_0\|.
\]
This implies that
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{1i} \leq \left( C_0 M \sqrt{n} \left\| \hat{\xi} - \xi_0 \right\|^2 \right) \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{\epsilon}_i \right)
\]
By assumption 4.4 and 4.5, the first term in the r.h.s. is an \( o_P(1) \). The second term is a \( O_P(1) \) by assumption 4.6 and the weak law of large numbers. The result follows.
where in the second inequality we use the fact that \( \hat{n} \) by assumption 4.4, 4.5, 4.2 and 4.7, hence, it suffices to prove that the second term tends to zero in

\[
\|R_2\| \leq \|S_i(\xi_0)\| \left[ \int_{T_i} \hat{q}(u, \hat{\xi})du - q_0(T_i, \xi_0)(\hat{T}_i - T_i) \right] \left[ 1\{\hat{T}_i \in \mathcal{S}\} + 1\{\hat{T}_i \notin \mathcal{S}\} \right]
\]

By (7.8), the right hand side tends to one. This establishes (7.10).

Moreover, by lemma 8.3, \( \sup_{u \in \mathcal{S}} |\hat{q}(u, \xi) - q_0(u, \xi_0)| = o_P(1) \). Now, let us prove that

\[
\max_{i, \hat{T}_i \in \mathcal{S}} |q_0(\hat{T}_i, \xi_0) - q_0(T_i, \xi_0)| = o_P(1) \tag{7.10}
\]

Fix \( \varepsilon > 0 \). Because \( q_0(\cdot, \xi_0) \) is continuous by assumption 4.2 and \( \mathcal{S} \) is compact, \( q_0(\cdot, \xi_0) \) is uniformly continuous on \( \mathcal{S} \). Thus, there exists \( \delta > 0 \) such that for all \( (u, v) \in \mathcal{S}^2 \) satisfying \( |u - v| \leq \delta \), we have \( |q_0(u, \xi_0) - q_0(v, \xi_0)| \leq \varepsilon \). As a consequence,

\[
P \left( \max_{i, \hat{T}_i \in \mathcal{S}} \left| q_0(\hat{T}_i, \xi_0) - q_0(T_i, \xi_0) \right| \leq \varepsilon \right) \geq P \left( \max_{i, \hat{T}_i \in \mathcal{S}} |\hat{T}_i - T_i| \leq \delta \right).
\]

By (7.8), the right hand side tends to one. This establishes (7.10). It remains to show that

\[
\left[ \sqrt{n} \|\hat{\xi} - \xi_0\| \right] \left[ \frac{1}{n} \sum_{i=1}^{n} \|S_i(\xi_0)\| 1\{\hat{T}_i \notin \mathcal{S}\} \right] = o_P(1) \tag{7.11}
\]

By assumptions 4.4 and 4.5, it suffices to prove that the second term tends to zero in
probability. By assumption 4.7 and the Cauchy-Schwartz inequality,

\[ E \left[ \|S_i(\xi_0)\| \, 1 \{\hat{T}_i \not\in S\} \right] \leq E \left[ \|S_i(\xi_0)\|^2 \right]^{1/2} P(\hat{T}_i \not\in S)^{1/2} \\
\leq E \left[ \|S_i(\xi_0)\|^2 \right]^{1/2} \left( P(\hat{T}_i > \overline{s}) + P(\hat{T}_i \leq \underline{s}) \right)^{1/2} \\
\leq E \left[ \|S_i(\xi_0)\|^2 \right]^{1/2} \left[ P \left( T_i + M \left\| \hat{\xi} - \xi_0 \right\| > \overline{s} \right) \\
+ P \left( T_i - M \left\| \hat{\xi} - \xi_0 \right\| < \underline{s} \right) \right]^{1/2} \tag{7.12} \]

where in the third inequality we use (7.8). Because \( T_i \in [\underline{s}, \overline{s}] \) and \( \left\| \hat{\xi} - \xi_0 \right\| \) tends to zero in probability, the r.h.s. of (7.12) tends to zero. Because convergence in \( L^1 \) implies convergence in probability, the second term of (7.11) tends to zero, proving \((\sum_{i=1}^{n} R_{2i})/\sqrt{n} = o_P(1)\).

– \( R_3 \): By the mean value theorem, there exists \( \tilde{\xi}_u \) in the segment between \( \xi_0 \) and \( \hat{\xi} \) such that

\[ \tilde{q}(u, \hat{\xi}) - \tilde{q}(u) = \frac{\partial \tilde{q}}{\partial \xi}(u, \tilde{\xi}_u) (\hat{\xi} - \xi_0) \]

Thus,

\[ |R_3| = \|S_i(\xi_0)\| \left\| \int_{\xi_0}^{\hat{T}_i} \frac{\partial \tilde{q}}{\partial \xi}(u, \tilde{\xi}) - \frac{\partial q_0}{\partial \xi}(u, \xi_0) \, du \right\| (\hat{\xi} - \xi_0) \]

\[ \leq C_2 \|S_i(\xi_0)\| \left\| \tilde{\xi} - \xi_0 \right\| \sup_{u \in S} \left\| \frac{\partial \tilde{q}}{\partial \xi}(u, \tilde{\xi}) - \frac{\partial q_0}{\partial \xi}(u, \xi_0) \right\| . \]

The supremum tends to zero in probability by lemma 8.3. The result follows by (7.9).

– \( R_4 \): this remainder term is similar to the one of the consumer surplus example of Newey & McFadden (1994, p. 2195 and 2204). The result follows by assumption 4.8, the rate condition on \( h_n \) and lemma 8.10 of Newey & McFadden (1994).

– \( R_5 \): first, note that

\[ \left| V_i(\hat{\lambda}) - V_i(\lambda_0) \right| = \left| X_i'(D_1(\beta_1 - \hat{\beta}_1) + (1 - D_1)(\beta_0 - \hat{\beta}_0)) + D_i X_i'(\xi_0 - \hat{\xi}) \\
+ \int_{T_i}^{\hat{T}_i} \tilde{q}(u, \hat{\xi}) \, du + \int_{\xi_0}^{T_i} \left[ \tilde{q}(u, \hat{\xi}) - q_0(u, \xi_0) \right] \, du \right| \]

\[ \leq M \left( \left\| \beta_1 - \beta_1 \right\| + \left\| \beta_0 - \beta_0 \right\| + 2 \left\| \hat{\xi} - \xi_0 \right\| + C_2 \sup_{u \in S} |\tilde{q}(u, \xi) - q_0(u, \xi_0)| \right) \]

Moreover, by assumption 4.7, there exists \( C_3 > 0 \) such that

\[ \left\| S_i(\hat{\xi}) - S_i(\xi_0) \right\| \leq C_3 \left\| \hat{\xi} - \xi_0 \right\|. \]
Thus,

$$
\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{5i} \right| \leq \left[ MC_3 \sqrt{n} \left\| \hat{\xi} - \xi_0 \right\| \right] \left( \left\| \hat{\beta}_1 - \beta_1 \right\| + \left\| \hat{\beta}_1 - \beta_1 \right\| \right)
$$

\quad + 2 \left\| \hat{\xi} - \xi_0 \right\| + C_2 \sup_{u \in S} |\tilde{q}(u, \xi) - q_0(u, \xi_0)| \right].
$$

By assumptions 4.4 and 4.5, the first term in the r.h.s. is an \( O_P(1) \). By lemma 8.3 and assumptions 4.4-4.5, the second term is an \( o_P(1) \). The result follows.

Step 2. Now, let us show the asymptotic normality of \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G(U_i, \tilde{\lambda} - \lambda_0) \). Let \( \alpha_0 = (\xi_0, \beta_1, \beta_0)' \) and \( \hat{\alpha} = (\hat{\xi}, \hat{\beta}_1, \hat{\beta}_0)' \). We have

$$
G(U_i, \tilde{\lambda} - \lambda_0) = P_i' (\hat{\alpha} - \alpha_0) + \tilde{G}(U_i, \tilde{r}, \tilde{f}),
$$

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

$$
P_{1i} = (V_i(\lambda_0) - W_i(\theta_0) \frac{\partial S_i}{\partial \xi} (\xi_0)' - S_i(\xi_0) \left( D_i X_i' + q_0(T_i, \xi_0) X_i' + \int_{T_i}^{T_0} \frac{\partial q_0}{\partial \xi}(u, \xi_0) du \right))
$$

$$
P_{2i} = -D_i S_i(\xi_0) X_i'
$$

$$
P_{3i} = -(1 - D_i) S_i(\xi_0) X_i'
$$

$$
\tilde{G}(U_i, \tilde{r}, \tilde{f}) = \int_{T_i}^{T_0} (1/f_0(u))(\tilde{r}(u) - q_0(u) \tilde{f}(u)) du.
$$

By the weak law of large numbers,

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

Moreover, we have \( \hat{\xi} = (\hat{\beta}_0m - \hat{\beta}_{1m})\hat{\xi} \). Thus, by assumptions 4.4 and 4.5,

$$
\hat{\xi} - \xi_0 = \frac{1}{n} \sum_{i=1}^{n} \tilde{\psi}_i + o_P \left( \frac{1}{\sqrt{n}} \right)
$$

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

$$
\tilde{\psi}_i = (\psi_{0mi} - \psi_{1mi})\hat{\xi} + (\hat{\beta}_0m - \hat{\beta}_{1m})\psi_i.
$$

Hence,

$$
\hat{\alpha} - \alpha_0 = \frac{1}{n} \sum_{i=1}^{n} (\tilde{\psi}_i, \psi_{1i}, \psi_{0i})' + o_P \left( \frac{1}{\sqrt{n}} \right).
$$

Thus,

$$
\frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} P_i \right)' (\hat{\alpha} - \alpha_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega_{1i} + o_P(1).
$$
where
\[ \Omega_{1i} = E[P_{1i}] \left( \tilde{\psi}_i, \psi_{1i}, \psi_{0i} \right)' \]  

(7.13)

Thus, it suffices to focus on the nonparametric part of \( G, \tilde{G}(U_i, r, f) \). Now, \( \tilde{G} \) is nearly the linearized part of the consumer surplus example of Newey & McFadden (1994, p. 2204), except that \( b \) is replaced by \( T_i \). Thus, it suffices to modify conveniently their proof (see Newey & McFadden, 1994, p. 2211), by checking conditions (ii), (iii) and (iv) as well as the technical requirements of their theorem 8.11. As a result, we get
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{G}(U_i, r, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega_{2i} + o_P(1)
\]

where \( \Omega_{2i} = S_T(T_i)1\{T_i \geq t_0\}(D_i - q_0(T_i))/f_0(T_i) \), with \( S_T(.) \) denoting the survival function of \( T \). The result follows.

Step 3. Eventually, we establish the asymptotic normality of \( \hat{\theta} \). By steps 1 and 2 and the central limit theorem applied on \( g(U_i, \lambda_0) \),
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(U_i, \hat{\lambda}) \rightarrow_{d} N \left( 0, V(g(U_1, \lambda_0) + \Omega_{11} + \Omega_{21}) \right).
\]

Thus, by definition of \( \hat{\theta} \) and \( g(U_i, \theta, \lambda) \),
\[
\left[ \frac{1}{n} \sum_{i=1}^{n} S_i(\hat{\xi})W_i' \right] \sqrt{n} (\hat{\theta} - \theta_0) \rightarrow_{d} N \left( 0, V(g(U_1, \lambda_0) + \Omega_{11} + \Omega_{21}) \right).
\]

Now, by assumption 4.7, \( \left\| S_i(\hat{\xi}) - S_i(\xi_0) \right\| \leq C_4 \left\| \hat{\xi} - \xi_0 \right\| \) for a given \( C_4 > 0 \). Thus,
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} S_i(\hat{\xi})W_i' - E(S_1(\xi_0)W_1') \right\| \leq C_4 \left( \frac{1}{n} \sum_{i=1}^{n} \left\| W_i \right\| \right) \left\| \hat{\xi} - \xi_0 \right\| + \frac{1}{n} \sum_{i=1}^{n} S_i(\xi_0)W_i' - E(S_1(\xi_0)W_1').
\]

Thus, by the weak law of large numbers,
\[
\frac{1}{n} \sum_{i=1}^{n} S_i(\hat{\xi})W_i' \overset{p}{\rightarrow} E(S_1(\xi_0)W_1') = E(S_1W_1').
\]

Eventually, by Slutski’s lemma, given that \( g(U_1, \lambda_0) = \eta S_1 \),
\[
\sqrt{n} (\hat{\theta} - \theta_0) \rightarrow_{d} N \left( 0, E(S_1W_1')^{-1}V(\eta S_1 + \Omega_{11} + \Omega_{21})E(W_1S_1')^{-1} \right)
\]

\( \square \)
Lemma 8.1  For all \( h \in I \) and all \( t < \lambda - G < t' \),

\[
\varphi_h(t', \lambda) - \varphi_h(t, \lambda) = \int \Delta P_{t,t',G,\lambda}(w) dh(w)
\]

where

\[
\Delta P_{t,t',G,\lambda}(w) = P (\varepsilon_0 \geq w, \varepsilon_0 + t + G \leq \varepsilon_1 \leq w + \lambda) - P (\varepsilon_0 \leq w, \varepsilon_0 + t' + G \geq \varepsilon_1 \geq w + \lambda)
\]

and, for any measurable function \( a \) and any function of bounded variation \( h \), \( \int a(u) dh(u) \) denotes the Lebesgue-Stieltjes integral.

Proof: by (3.3), we have

\[
\varphi_h(t', \lambda) - \varphi_h(t, \lambda) = \int_t^{t'} \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) [h(v) - h(v + G - \lambda + u)] dvdu
\]

\[
= \int_t^{\lambda - G} \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) [h(v) - h(v + G - \lambda + u)] dvdu
\]

\[
- \int_{\lambda - G}^{t'} \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) [h(v + G - \lambda + u) - h(v)] dvdu.
\]

Let us focus on the first integral. By definition of the Lebesgue-Stieltjes integral,

\[
h(v) - h(v + G - \lambda + u) = \int 1\{v + G - \lambda + u \leq w \leq v\} dh(w).
\]

Hence, by Fubini’s theorem on nonnegative function, we get

\[
\int_t^{\lambda - G} \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) [h(v) - h(v + G - \lambda + u)] dvdu
\]

\[
= \int_t^{\lambda - G} \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) \left[ \int 1\{v + G - \lambda + u \leq w \leq v\} dh(w) \right] dvdu
\]

\[
= \int \left[ \int \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) 1\{w \leq v\} 1\{v + G + t \leq v + G + u \leq w + \lambda\} dvdu \right] dh(w)
\]

\[
= \int P (\varepsilon_0 \geq w, \varepsilon_0 + t + G \leq \varepsilon_1 \leq w + \lambda) dh(w).
\]

Similarly,

\[
\int_{\lambda - G}^{t'} \int f_{\varepsilon_0, \varepsilon_1}(v, v + G + u) [h(v + G - \lambda + u) - h(v)] dvdu
\]

\[
= \int P (\varepsilon_0 \leq w, \varepsilon_0 + t' + G \geq \varepsilon_1 \geq w + \lambda) dh(w).
\]

The lemma follows.
Lemma 8.2 Suppose that assumptions 4.2 and 4.6 hold. Then, for all \(u \in \mathcal{S}, \xi \mapsto f_0(u, \xi)\) and \(\xi \mapsto r_0(u, \xi)\) admit partial derivatives at \(\xi_0\) which satisfy:

\[
\frac{\partial f_0}{\partial \xi}(u, \xi_0) = -(E[X|T = u] f_T(u))' \tag{8.1}
\]

\[
\frac{\partial r_0}{\partial \xi}(u, \xi_0) = -(E[DX|T = u] f_T(u))' \tag{8.2}
\]

Proof: let \(X_{-m} = (X_1, ..., X_{m-1}, X_{m+1}, ..., X_p)\) and \(f_{X_m|X_{-m}}(., x)\) (resp. \(F_{X_m|X_{-m}}(., x)\)) denote the density (resp cdf) 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

\[
f_0(u, \xi + t\delta_k) = \begin{cases} 
E \left[ f_{X_m|X_{-m}} \left( \frac{u - X_{m, \xi - 1} - tX_k}{\xi}, X_{-m} \right) \right] & \text{if } k \neq m \\
E \left[ f_{X_m|X_{-m}} \left( \frac{u - X_{m, \xi - 1}}{\xi}, X_{-m} \right) \right] & \text{if } k = m 
\end{cases}
\]

Thus, by assumption 4.2 and dominated convergence, \(\xi \mapsto f_0(u, \xi)\) admits continuous partial derivatives. Now, let \(F_0(., \xi)\) denote the cdf of \(X'\xi\). We have, for \(k > 1\),

\[
F_0(u, \xi + t\delta_k) = \begin{cases} 
E \left[ F_{X_m|X_{-m}} \left( \frac{u - X_{m, \xi - 1} - tX_k}{\xi}, X_{-m} \right) \right] & \text{if } k \neq m \\
E \left[ F_{X_m|X_{-m}} \left( \frac{u - X_{m, \xi - 1}}{\xi}, X_{-m} \right) \right] & \text{if } k = m 
\end{cases}
\]

Thus, by assumption 4.2 and dominated convergence, \(\xi \mapsto F_0(u, \xi)\) admits continuous partial derivatives, and after some rearrangements,

\[
\frac{\partial F_0}{\partial \xi_k}(u, \xi_0) = -E[X_k|T = u] f_0(u, \xi_0).
\]

By assumption 4.2 once more, \(u \mapsto \partial F_0/\partial \xi_k(u, \xi_0)\) is continuously differentiable and

\[
\frac{\partial^2 F_0}{\partial u \partial \xi}(u, \xi_0) = -(E[X|T = u] f_0(u, \xi_0))'.
\]

Then (8.1) follows from \(\partial f_0/\partial \xi = \partial^2 F_0/\partial \xi \partial u = \partial^2 F_0/\partial u \partial \xi\).

The proof of (8.2) is similar, except that we use \(G_0(u, \xi) = E(D1\{X'\xi \leq u\})\) instead of \(F_0(u, \xi)\). The partial derivatives of \(\xi \mapsto G_0(u, \xi)\) exist and satisfy

\[
\frac{\partial G_0}{\partial \xi}(u, \xi) = -E(DX|T = u) f_0(u, \xi_0)
\]

\[
= -F_{\epsilon_0 - \epsilon_1}(-u - \delta)E(X|T = u) f_0(u, \xi_0).
\]

Then differentiability of \(u \mapsto \partial G_0/\partial \xi(u, \xi)\) stems from assumptions 4.2 and 4.6. (8.2) follows by the same argument as previously.
Lemma 8.3 Suppose that \( nh_n^6 \to \infty \), \( nh_n^8 \to 0 \) and assumptions 4.2, 4.8 hold. Then, for all \( \xi_n \) such that \( \| \xi_n - \xi_0 \| = O_P(1/\sqrt{n}) \), we have

\[
\sup_{u \in S} |\hat{q}(u, \xi_n) - q_0(u, \xi_0)| = o_P(1) \tag{8.3}
\]

\[
\sup_{u \in S} \left\| \frac{\partial \hat{q}}{\partial \xi}(u, \xi_n) - \frac{\partial q_0}{\partial \xi}(u, \xi_0) \right\| = o_P(1) \tag{8.4}
\]

Proof: we first write

\[
\sup_{u \in S} |\hat{q}(u, \xi_n) - q_0(u, \xi_0)| \leq \sup_{u \in S} |\hat{q}(u, \xi_n) - \hat{q}(u, \xi_0)| + \sup_{u \in S} |\hat{q}(u, \xi_0) - q_0(u, \xi_0)| \tag{8.5}
\]

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

\[
\sup_{u \in S} |\hat{q}(u, \xi_n) - \hat{q}(u, \xi_0)| = \sup_{u \in S} \frac{1}{\hat{f}(u, \xi_0)} \left| (\hat{r}(u, \xi_n) - \hat{r}(u, \xi_0)) + \hat{q}(u, \xi_n)(\hat{f}(u, \xi_0) - \hat{f}(u, \xi_n)) \right|
\]

\[
\leq \sup_{u \in S} \frac{1}{\hat{f}(u, \xi_0)} \left[ |\hat{r}(u, \xi_n) - \hat{r}(u, \xi_0)| + |\hat{f}(u, \xi_n) - \hat{f}(u, \xi_0)| \right]
\]

\[
\leq \sup_{u \in S} |\hat{r}(u, \xi_n) - \hat{r}(u, \xi_0)| + \sup_{u \in S} |\hat{f}(u, \xi_n) - \hat{f}(u, \xi_0)| \tag{8.6}
\]

Let us prove that

\[
\sup_{u \in S} \left| \hat{f}(u, \xi_n) - \hat{f}(u, \xi_0) \right| = o_P(1) \tag{8.7}
\]

The proof for \( \hat{r} \) is similar. By assumption 4.8, there exists \( C_5 > 0 \) such that \( |K(u) - K(v)| \leq C_5 |u - v| \). Thus,

\[
\left| \hat{f}(u, \xi_n) - \hat{f}(u, \xi_0) \right| \leq \frac{1}{nh_n} \sum_{i=1}^{n} \left| K \left( \frac{u - X_i' \xi_n}{h_n} \right) - K \left( \frac{u - X_i' \xi_0}{h_n} \right) \right|
\]

\[
\leq \frac{C_5 M \| \xi_n - \xi_0 \|}{h_n^2} = O_p \left( \frac{1}{\sqrt{nh_n^2}} \right).
\]

This establishes (8.7) since \( nh_n^4 \to \infty \). This also proves that \( \inf_{u \in S} \hat{f}(u, \xi_0) \) converges in probability to \( \inf_{u \in S} f_0(u, \xi_0) \), which is positive by assumption 4.2. 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 4.2 and 4.8, and conditions on \( h_n \) ensure that we can apply lemma 8.10 of Newey & McFadden (1994), yielding \( \sup_{u \in S} |\hat{f}(u, \xi_0) - f(u, \xi_0)| = o_P(1) \) and similarly for \( \hat{r}(., \xi_0) \). This establishes (8.3).
Now, let us turn to (8.4). We use the same decomposition as (8.5). First, let us establish that
\[ \sup_{u \in S} \left| \frac{\partial \hat{q}}{\partial \xi}(u, \xi_0) - \frac{\partial q_0}{\partial \xi}(u, \xi_0) \right| = O_P(1) \]  
(8.8)

We have
\[ \frac{\partial \hat{q}}{\partial \xi}(u, \xi_0) = \frac{1}{f(u, \xi_0)} \left[ \frac{\partial \hat{r}}{\partial \xi}(u, \xi_0) - \hat{q}(u, \xi_0) \frac{\partial \hat{f}}{\partial \xi}(u, \xi_0) \right]. \]
and similarly for \( \frac{\partial q_0}{\partial \xi}(u, \xi_0) \). Thus,
\[ \frac{\partial \hat{q}}{\partial \xi}(u, \xi_0) - \frac{\partial q_0}{\partial \xi}(u, \xi_0) \]
\[ = \frac{1}{f(u, \xi_0)} \left\{ \left[ \frac{\partial \hat{r}}{\partial \xi}(u, \xi_0) - \frac{\partial r_0}{\partial \xi}(u, \xi_0) \right] - \frac{\partial r_0}{\partial \xi}(u, \xi_0) \left[ \frac{\hat{f}(u, \xi_0) - f_0(u, \xi_0)}{f_0(u, \xi_0)} \right] \right\} \]
\[ - \frac{\hat{q}(u, \xi_0)}{f(u, \xi_0)} \left[ \left( \frac{\partial \hat{f}}{\partial \xi}(u, \xi_0) - \frac{\partial f_0}{\partial \xi}(u, \xi_0) \right) - \frac{\partial f_0/\partial \xi(u, \xi_0)}{f_0(u, \xi_0)} \left( \hat{f}(u, \xi_0) - f_0(u, \xi_0) \right) \right] \]
\[ - \frac{\partial f_0/\partial \xi(u, \xi_0)}{f_0(u, \xi_0)} (\hat{q}(u, \xi_0) - q_0(u, \xi_0)) \]

By what precedes, \( \inf_{u \in S} \hat{f}(u, \xi_0) \) tends in probability to \( \inf_{u \in S} f(u, \xi_0) > 0 \), while \( \sup_{u \in S} |\hat{f}(u, \xi_0) - f_0(u, \xi_0)| = O_P(1) \). Besides, \( \hat{q}(., \xi_0) \) is bounded by 1 and by lemma 8.2, \( \partial f_0/\partial \xi(., \xi_0) \) is continuous on the compact set \( S \) and thus is bounded on this set. Thus, it suffices to prove that
\[ \sup_{u \in S} \left| \frac{\partial \hat{f}}{\partial \xi}(u, \xi_0) - \frac{\partial f_0}{\partial \xi}(u, \xi_0) \right| = O_P(1) \]  
(8.9)
and similarly for \( r_0 \). By lemma 8.2, \( u \mapsto \partial f_0/\partial \xi(u, \xi_0) \) is the derivative of \( -E(X|T = u) f_0(u) \). As a consequence, we can apply Newey & McFadden (1994)’s lemma 8.10, using as before assumptions 4.2, 4.8, and conditions on \( h_n \). This yields (8.9). The same reasoning applies to \( r_0 \), showing (8.8).

Now, let us establish that
\[ \sup_{u \in S} \left\| \frac{\partial \hat{q}}{\partial \xi}(u, \xi_n) - \frac{\partial \hat{q}}{\partial \xi}(u, \xi_0) \right\| = O_P(1) \]
Using a similar decomposition as previously and the preceding results, it suffices to prove that
\[ \sup_{u \in S} \left\| \frac{\partial \hat{f}}{\partial \xi}(u, \xi_n) - \frac{\partial \hat{f}}{\partial \xi}(u, \xi_0) \right\| = O_P(1) \]  
(8.10)
and similarly for \( \hat{r} \). By assumption 4.8, there exists \( C_6 > 0 \) such that \(|K'(u) - K'(v)| \leq C_6|u - v| \). Thus,

\[
\left\| \frac{\partial \hat{f}}{\partial \xi}(u, \xi_n) - \frac{\partial \hat{f}}{\partial \xi}(u, \xi_0) \right\| \leq \frac{1}{nh_n^2} \sum_{i=1}^{n} \|X_i\| \left| K'\left( \frac{u - X_i' \xi_n}{h_n} \right) - K'\left( \frac{u - X_i' \xi_0}{h_n} \right) \right| \\
\leq \frac{C_6 M^2}{h_n^3} \frac{\|\xi_n - \xi_0\|}{h_n^3} = O_p\left( \frac{1}{\sqrt{nh_n^3}} \right).
\]

This proves (8.10) since \( nh_n^6 \rightarrow 0 \). The same reasoning applies to \( \hat{r} \). The result follows.
References


