Econometric Theory, **29**, 2013, 213–224.
doi:10.1017/S026646661200028X

ANOTHER LOOK AT THE IDENTIFICATION AT INFINITY OF SAMPLE SELECTION MODELS

XAVIER D'HAULTFOEUILLE
CREST

ARNAUD MAUREL
Duke University

It is often believed that without instruments, endogenous sample selection models are identified only if a covariate with a large support is available (see, e.g., Chamberlain, 1986, *Journal of Econometrics* 32, 189–218; Lewbel, 2007, *Journal of Econometrics* 141, 777–806) . We propose a new identification strategy mainly based on the condition that the selection variable becomes independent of the covariates for large values of the outcome. No large support on the covariates is required. Moreover, we prove that this condition is testable. We finally show that our strategy can be applied to the identification of generalized Roy models.

1. INTRODUCTION

Since the seminal work of Heckman (1974), the issue of endogenous selection has been an active topic of research in both applied and theoretical econometrics (for a survey, see Vella, 1998). The usual strategy to deal with this issue is to rely on instruments that determine selection but not the outcome. However, the search of a valid instrument may be difficult if not impossible in some applications. Another strategy, which has been sometimes advocated, relies on the fact that, loosely speaking, the selection problem becomes negligible "at the limit." Following this idea, Chamberlain (1986) proved that the effects of covariates on an outcome are identified under the linearity of the model and a large support assumption on at least one covariate. Lewbel (2007) generalized this result by proving that identification can be achieved without imposing any structure on the outcome equation, provided that a special regressor has a large support and under restrictions on the selection equation.[1]

The main drawback of the latter approach is that it requires the existence of a covariate with a large support. Thus, it breaks down when all covariates are

We are grateful to Magali Beffy, Edwin Leuven, and Arthur Lewbel for helpful comments. We also thank the editor Yuichi Kitamura and two anonymous referees for their valuable remarks and suggestions. Address correspondence to Xavier D'Haultfoeuille, CREST, 15 boulevard Gabriel Péri, 92 240 Malakoff, France; e-mail: xavier.dhaultfoeuille@ensae.fr.

© Cambridge University Press 2012

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

discrete, a case that is fairly common in practice. In this paper, we consider another route for identifying the model at the limit. Intuitively, if selection is truly endogenous, then we can expect the effect of the outcome on selection to dominate those of the covariates for large values of the outcome. Following this idea, our main identifying condition states that the selection variable is independent of the covariates at the limit, i.e., when the outcome tends to its upper bound. Under this condition, the model is identified without any large support condition on these covariates. Only an exogeneity assumption and a mild restriction on the residuals are required. Moreover, we show that the main condition is testable. Apart from the standard selection model, we apply our result to a generalization of the Roy (1951) model of self-selection accounting for nonpecuniary factors. In this framework, the effects of covariates on the outcomes are identified without exclusion restrictions under a moderate dependence condition on the residuals.

The paper is organized as follows. Section 2 presents the model and establishes the main identification result. Section 3 proves the testability of our main condition. Section 4 applies this result to generalized Roy models, and Section 5 concludes.

2. MAIN RESULT

Let Y^* denote the outcome of interest, X denote a vector of covariates, and D denote the selection dummy. Let us consider the following model, with $\sigma(X) > 0$:

$$Y^* = \psi(X) + \sigma(X)\varepsilon. \tag{2.1}$$

The econometrician observes D, $Y = DY^*$ and X. Without loss of generality, we suppose that $\psi(x_0) = 0$ and $\sigma(x_0) = 1$ for a given $x_0 \in \text{Supp}(X)$, where $\text{Supp}(T)$ denotes the support of the random variable T.[2] Our main result is based on the following assumptions.

Assumption 1 (Exogeneity)**.** $X \perp\!\!\!\perp \varepsilon$.

Assumption 2 (Restrictions on the tail of the residual)**.** Either $M = \sup(\text{Supp}(\varepsilon)) = \infty$, and there exists $\beta > 0$ such that $\mathbb{E}(\exp(\beta\varepsilon)) < \infty$, or $M < \infty$, and there exists $\gamma > 0$ such that $\mathbb{E}[1/(M-\varepsilon)^{\gamma}] < \infty$.

Assumption 3 (Independence at the limit)**.** There exists $l > 0$ such that for all $x \in \text{Supp}(X)$, $\lim_{y \to \overline{y}_x} P(D = 1 | X = x, Y^* = y) = l$, where $\overline{y}_x$ denotes the upper bound of the support of Y^* conditional on $X = x$.

Assumption 1 is usual in selection models and weaker than the exogeneity assumption imposed by Chamberlain (1986), because heteroskedasticity is allowed for here. Assumption 2 puts some restrictions on the tail of the distribution of ε.[3] In the example of a wage equation where Y^* denotes the logarithm of the wage W and $M = \infty$, it is satisfied if $\mathbb{E}[W^{\beta}] < \infty$ for a given $\beta > 0$. Thus,

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

it holds even if wages have very fat tails, e.g., Pareto-like. In standard examples, the support of Y^* is infinite, but Assumption 2 also accommodates a finite upper bound for ε, under a mild restriction. This restriction only rules out the existence of a mass point at the upper bound or a density tending to infinity very quickly at M (typically, a density equivalent to $1/\left[(M-x)\ln^2(M-x)\right]$).

Finally, Assumption 3 is the main condition here. It requires the probability of selection to be independent of X at the limit, i.e., for those who have large outcomes. In other terms, the effect of Y^* on selection becomes prominent when Y^* tends to its upper bound. To illustrate Assumption 3, let us consider the following selection rule:

$$D = \mathbb{1}\{\varphi(X) + \eta \geq 0\}. \tag{2.2}$$

Endogenous selection stems from the correlation between η and ε. Suppose that the following decomposition holds:

$$\eta = h(\varepsilon) + \nu, \qquad \nu \perp\!\!\!\perp (\varepsilon, X).$$

Then we get

$$D = \mathbb{1}\left\{\varphi(X) + h\left(\frac{Y^* - \psi(X)}{\sigma(X)}\right) + \nu \geq 0\right\}.$$

Thus, Assumption 3 is satisfied (with $l = 1$) provided that $h(x) \to \infty$ as $x \to \infty$.[4] In particular, when $h(x) = ax$, this condition holds provided that $a > 0$. Hence, in the Gaussian case, Assumption 3 is satisfied as soon as $\mathrm{Cov}(\eta, \varepsilon) > 0$. On the other hand, it fails to hold when $a = 0$ (unless $\varphi(\cdot)$ is constant). This is logical because this case corresponds to an exogenous selection where $P(D = 1|X = x, Y^* = y)$ is independent of y. As shown in Section 3, it is actually possible to reject Assumption 3 from the data in this case. It also fails to hold when $a < 0$, that is to say when ε and η are negatively correlated. In this case, however, Assumption 3 holds for small outcomes, by replacing Y^* by $-Y^*$. Thus, we can still apply Theorem 2.1, which follows, provided that $-\varepsilon$ satisfies the restrictions of Assumption 2.[5]

In the preceding examples, $l = 1$, but Assumption 3 also holds with $0 < l < 1$. This is the case (under the assumption that $\lim_{x\to\infty} h(x) = \infty$) if $D = U\mathbb{1}\{\varphi(X) + \eta \geq 0\}$, where $U \in \{0, 1\}$ is a random shock independent of (X, ε, η) satisfying $P(U = 1) > 0$. For instance, this framework may be used to model participation in the labor market, with U denoting in that case an unobserved random shock related to, e.g., health conditions that could prevent individuals from entering the labor market.

THEOREM 2.1. *Under Assumptions 1–3, $\psi(\cdot)$ and $\sigma(\cdot)$ are identified.*

Proof. Subsequently, S_T denotes the survival function of the random variable T. Besides, we use the notation $f(y) \sim g(y)$ if there exists $r(\cdot)$ such that

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

$f(y) = g(y)(1+r(y))$ with $\lim_{y\to\infty} r(y) = 0$. The result is based on the following lemma.

LEMMA 2.1. *Let T be a real random variable such that* $\sup(\text{Supp}(T)) = \infty$ *and* $\mathbb{E}(|T|) < \infty$. *Suppose also that* $S_T(y) \sim S_T(lf(y))$, *where* $\lim_{y\to\infty} f'(y) = 1$ *and* $l > 0$. *Then* $l = 1$.

Proof of Lemma 2.1. Suppose that $l > 1$. Then there exists $\eta > 0$ such that $l > 1+\eta$. Moreover, because $\sup(\text{Supp}(T)) = \infty$, $S_T(lf(y)) > 0$ for all y. Thus, $S_T(y) \sim S_T(lf(y))$ implies that there exists y_0 such that for all $y \geq y_0$,

$$S_T(y) < (1+\eta) S_T(lf(y)).$$

Besides, $\mathbb{E}(|T|) < \infty$ implies that $\int_0^\infty S_T(u)du < \infty$. Consequently, for all $y \geq y_0$,

$$\int_y^\infty S_T(u)du < (1+\eta)\int_y^\infty S_T(lf(u))du. \tag{2.3}$$

By assumption, the derivative of the function $m(y) = lf(y)$ tends to $l > 1$ when $y \to \infty$. Thus, there exists y_1 such that for all $y \geq y_1$, $m'(y) > 1+\eta$. Integrating between y_1 and $y \geq y_1$ shows that $m(y) > (1+\eta)(y - y_1) + m(y_1)$. Thus, there exists $y_2 \geq y_1$ such that $m(y) > y$ for all $y \geq y_2$. Hence, for all $y \geq y_2$, m is one-to-one and

$$\begin{aligned}\int_y^\infty S_T(lf(u))du &= \int_{m(y)}^\infty \frac{S_T(v)}{m'(m^{-1}(v))}dv \\ &< \frac{1}{1+\eta}\int_{m(y)}^\infty S_T(v)dv \\ &< \frac{1}{1+\eta}\int_y^\infty S_T(v)dv. \end{aligned} \tag{2.4}$$

Inequalities (2.3) and (2.4) imply that $\int_y^\infty S_T(u)du < \int_y^\infty S_T(u)du$ for all $y \geq \max(y_0, y_2)$, a contradiction. Similarly, one can show that $l < 1$ is impossible. Thus $l = 1$. ■

Now let us prove Theorem 2.1. First suppose that $\overline{y}_x = \infty$, or, equivalently, $\sup(\text{Supp}(\varepsilon)) = \infty$. Let $q(y,x) = P\left(D = 1, Y^* \geq y | X = x\right)$. We have

$$q(y,x) = \int_y^\infty P(D = 1|X = x, Y^* = u) dP^{Y^*|X=x}(u).$$

By Assumption 3, as $u \to \infty$, we have $P(D = 1|X = x, Y^* = u) \to l > 0$. Thus, using standard results on integrals, we get as $y \to \infty$,

$$q(y,x) \sim l\, P(Y^* \geq y | X = x).$$

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

By Assumption 1, $P(Y^* \geq y|X = x) = S_\varepsilon((y - \psi(x))/\sigma(x))$, where $S_\varepsilon(\cdot)$ denotes the survival function of ε. Thus,

$$q(y,x) \sim l S_\varepsilon\left(\frac{y-\psi(x)}{\sigma(x)}\right). \tag{2.5}$$

Similarly,

$$q(y,x_0) \sim l S_\varepsilon(y). \tag{2.6}$$

This implies that

$$q(y,x) \sim q\left(\frac{y-\psi(x)}{\sigma(x)}, x_0\right). \tag{2.7}$$

The function q is identified. Thus, $\sigma(x)$ and $\psi(x)$ are identified if, as $y \to \infty$,

$$q(y,x) \sim q(sy+u,x_0)\,(s>0) \implies (s,u) = \left(\frac{1}{\sigma(x)}, -\frac{\psi(x)}{\sigma(x)}\right). \tag{2.8}$$

To prove (2.8), suppose that $s > 0$ and u satisfy $q(y,x) \sim q(sy+u,x_0)$. Then it follows from (2.5) and (2.6) that

$$S_\varepsilon(t(y+v)) \sim S_\varepsilon(y), \tag{2.9}$$

where $t = s\sigma(x)$ and $v = (1/\sigma(x))(\psi(x) + u/s)$. Besides, by Assumption 2, $\mathbb{E}(|\varepsilon|) < \infty$. Thus, by Lemma 2.1, $t = 1$; i.e., $s = 1/\sigma(x)$. Thus, $\sigma(x)$ is identified. Besides, by (2.9),

$$S_{e^{\beta\varepsilon}}(wy) \sim S_{e^{\beta\varepsilon}}(y),$$

where β is defined in Assumption 2 and $w = \exp(\beta v)$. Because $\mathbb{E}(\exp(\beta\varepsilon)) < \infty$, we can apply Lemma 2.1 once more. This yields $w = 1$, which is equivalent to $u = -\psi(x)/\sigma(x)$. Thus, $\psi(x)$ is identified.

Now, let us turn to the case where $\overline{y}_x < \infty$. Instead of $q(y,x)$, consider $r(y,x) = P(D = 1, Y^* \geq \overline{y}_x - (1/y^{1/\gamma})\,|\,X = x)$ and let $T = 1/(M-\varepsilon)^\gamma$. Reasoning as previously, we have, as $y \to \infty$,

$$r(y,x) \sim l\, S_T\left(\sigma(x)^\gamma\, y\right).$$

Thus, $r(y,x) \sim r(\sigma(x)^\gamma\, y, x_0)$, and $\sigma(x)$ is identified if $S_T(uy) \sim S_T(y)$ implies that $u = 1$. This is the case by Assumption 2 and Lemma 2.1. It follows from Assumption 3 that $M = \overline{y}_{x_0}$ and $\overline{y}_x$ are identified, thus implying that $\psi(x) = \overline{y}_x - \sigma(x)M$ is also identified. ■

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

The key point of the proof is that by Assumption 3, the conditional survival function of Y is equivalent (up to a constant) to the one of a location-scale model. Then the normalization $(\psi(x_0), \sigma(x_0)) = (0, 1)$ and the restrictions on ε ensure that the parameters of this location-scale model can be identified. Unlike Lewbel (2007), we impose additive separability in the outcome equation. On the other hand, no structure is imposed on the selection process, apart from Assumption 3.

By assuming independence at the limit between the probability of selection and the covariates, we are able to identify the covariate effects on the potential outcome directly from the observed effects on the conditional survival function of Y. This idea is closely related to the identification strategy developed for mixed proportional hazards models (for a recent review, see, e.g., Abbring, 2010).[6] In these models, identification relies on the fact that, for durations close to zero, the survival outcome is observed for the whole population, irrespective of the covariate values (see Elbers and Ridder, 1982). As a result, the covariate effects on the hazard rate, conditional on the covariates and unobserved heterogeneity, can be identified at the limit from the observed effects on the mean hazard rate, conditional on the covariates only.[7] As in our case, this approach does not require a large support condition on the covariates. Overall, this sheds an interesting light on the connection between static and dynamic selection bias issues.

Theorem 2.1 does not provide any information on the intercept of (2.1), i.e., on $\mathbb{E}(\varepsilon)$. Actually, one can show that this intercept is not identified in general in our context. Basically, this stems from the fact that contrary to the framework of Heckman (1990) or Andrews and Schafgans (1998), there is in general no individual for whom $P(D = 1|X)$ is arbitrarily close to one. Moreover, apart from Assumption 3, our model puts no restriction on the probability $P(D = 1|X, Y^*)$. As a result, it is possible to define a distribution for ε and a conditional probability of selection different from the true ones but observationally equivalent, leading to different values for $\mathbb{E}(\varepsilon)$.

3. TESTABILITY

The main identifying condition in the preceding setting is Assumption 3, and so one may wonder whether this assumption is refutable or not. The answer turns out to be affirmative. To see this, note that this condition, together with Assumptions 1 and 2, implies (2.7), which can be stated as[8]

$$\forall x \in \text{Supp}(X),\ \exists (s(x), u(x)) \in \mathbb{R}^{*+} \times \mathbb{R} :\ q(y, x) \sim q(s(x)y + u(x), x_0), \quad \textbf{(3.1)}$$

where $q(y, x) = P(D = 1, Y^* \geq y|X = x)$. Because the function q is identified, condition (3.1) can be tested in the data. Then one can reject Assumption 3 when there is no $(s(x), u(x))$ satisfying (3.1). Theorem 3.1, which follows, shows that the reverse also holds: under a slight reinforcement of Assumption 2 and another mild condition, condition (3.1) and Assumption 3 are equivalent. This means that we can reject Assumption 3 whenever it fails to hold.

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

THEOREM 3.1. *Suppose that Assumption 1 holds,* $\sup(\mathrm{Supp}(\varepsilon)) = \infty$, *there exists* $\alpha > 1$, $\beta > 0$ *such that* $\mathbb{E}[\exp(\beta|\varepsilon|^{\alpha})] < \infty$, *and there exists* $l(x) > 0$ *such that*

$$\lim_{y\to\infty} P(D = 1|X = x, Y^* = y) = l(x). \tag{3.2}$$

Then Assumption 3 is equivalent to condition (3.1).

Proof. We shall first prove a result similar to the one of Lemma 2.1.

LEMMA 3.1. *Let* T *be a real random variable such that* $\sup(\mathrm{Supp}(T)) = \infty$ *and* $\mathbb{E}(|T|) < \infty$. *Suppose also that when* $y \to \infty$, $S_T(y) \sim l S_T(f_\delta(y))$, *where* $l > 0$ *and* $f_\delta(\cdot)$ *is strictly increasing for* y *large enough and satisfies*

(i) $f'_\delta(y) \to 0$ *if* $\delta < 0$,
(ii) $f'_0(y) \to C > 0$, *and*
(iii) $f'_\delta(y) \to \infty$ *if* $\delta > 0$.

Then $\delta = 0$. *Moreover, if* $f_0(y) = y$, *then* $l = 1$.

Proof of Lemma 3.1. Suppose that $\delta > 0$. By assumption, there exist $l' > 0$ and y_0 such that for all $y \geq y_0$,

$$S_T(y) < l' S_T(f_\delta(y)). \tag{3.3}$$

Besides, there exists y_1 such that $f_\delta(\cdot)$ is one-to-one on $[y_1, \infty)$, with $f'_\delta(y) > l'$ and $f_\delta(y) > y$ for all $y \geq y_1$. Thus, for all $y \geq y_1$,

$$\begin{aligned}\int_y^\infty S_T(f_\delta(u))du &= \int_{f_\delta(y)}^\infty \frac{S_T(v)}{f'_\delta(f_\delta^{-1}(v))}dv \\ &< \frac{1}{l'}\int_{f_\delta(y)}^\infty S_T(v)dv \\ &< \frac{1}{l'}\int_y^\infty S_T(v)dv. \end{aligned} \tag{3.4}$$

Inequalities (3.3) and (3.4) imply that $\int_y^\infty S_T(u)du < \int_y^\infty S_T(u)du$ for all $y \geq \max(y_0, y_1)$, a contradiction. The proof that $\delta < 0$ is impossible follows similarly. Thus $\delta = 0$. Finally, if $f_0(y) = y$, then $S_T(y) \sim l S_T(y)$, which implies directly that $l = 1$. ■

Now let us prove Theorem 3.1. By the proof of Theorem 2.1, Assumption 3 implies condition (3.1). Thus, it suffices to prove that condition (3.1) implies Assumption 3. For all $x \in \mathrm{Supp}(X)$, by similar reasoning as in the previous proof,

$$q(y, x) \sim l(x) S_\varepsilon\left(\frac{y - \psi(x)}{\sigma(x)}\right).$$

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

The same holds for $q(y, x_0)$. Thus, by condition (3.1), there exist $\mu > 0$ and $\nu \in \mathbb{R}$ such that

$$S_\varepsilon(y) \sim l S_\varepsilon(\mu y + \nu), \tag{3.5}$$

where $l = l(x)/l(x_0)$. This implies that

$$S_{\exp(\beta\varepsilon)}(y) \sim l S_{\exp(\beta\varepsilon)}(\exp(\beta\nu) y^\mu).$$

By assumption, $\mathbb{E}[\exp(\beta\varepsilon)] < \infty$. Thus, by applying Lemma 3.1 to $f_\delta(y) = \exp(\beta\nu) y^{\exp(\delta)}$ (with $\delta = \ln \mu$), we get $\mu = 1$. Hence, by (3.5),

$$S_{\exp(\beta\varepsilon^\alpha)}(\exp(\beta y^\alpha)) \sim l S_{\exp(\beta\varepsilon^\alpha)}(\exp(\beta(y+\nu)^\alpha)).$$

After some manipulations, we obtain

$$S_{\exp(\beta\varepsilon^\alpha)}(y) \sim l S_{\exp(\beta\varepsilon^\alpha)}(f_\nu(y)),$$

where

$$f_\nu(y) = y^{\left(1+\nu(\beta/\ln y)^{1/\alpha}\right)^\alpha}.$$

Some computations show that f_ν is strictly increasing for y large enough and

(a) $f_\nu'(y) \to 0$ if $\nu < 0$,
(b) $f_0(y) = y$, and
(c) $f_\nu'(y) \to \infty$ if $\nu > 0$.

Thus, by Lemma 3.1, $\nu = 0$ and $l = 1$. In other terms, $l(x) = l(x_0)$ for all $x \in \text{Supp}(X)$, which proves that Assumption 3 holds. ■

To illustrate Theorem 3.1, suppose, e.g., that in the true model, selection is exogenous, i.e., $P(D = 1|X = x, Y^* = y) = P(D = 1|X = x)$ for all y, and that $x \mapsto P(D = 1|X = x)$ is nonconstant, so that Assumption 3 fails to hold. In this setting, condition (3.2) is satisfied with $l(x) = P(D = 1|X = x)$. Thus, by Theorem 3.1, condition (3.1) fails to hold. This means that the "independence at the limit" assumption can be rejected by the data when selection is exogenous. Theorem 3.1 is also useful when one does not know a priori whether Assumption 3 holds for large or small values of the outcome. Indeed, as noted before, the limit of $P(D = 1|X = x, Y^* = y)$ may be independent of x when y tends to $-\infty$ rather than $+\infty$, in which case identification is still achieved. Under the conditions stated in Theorem 3.1, and if $\inf(\text{Supp}(\varepsilon)) = -\infty$ and $\lim_{y\to-\infty} P(D = 1|X = x, Y^* = y)$ exists and is strictly positive, the result of the theorem holds at both $-\infty$ and $+\infty$, so that one can test for the two conditions and choose the appropriate restriction.

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

4. APPLICATION TO GENERALIZED ROY MODELS

We consider a class of generalized Roy models where each individual chooses the sector $D \in \{0, 1\}$ that provides her with the higher utility. Suppose that the utility U_i associated with each sector $i \in \{0, 1\}$ is the sum of the potential log-earnings $Y_i = \psi_i(X) + \varepsilon_i$ and a random nonpecuniary component $G_i(X) + \eta_i$. Thus, $D = \mathbb{1}\{Y_1 \geq Y_0 + G(X) + \eta\}$ with $G(X) = G_0(X) - G_1(X)$ and $\eta = \eta_0 - \eta_1$, and the econometrician only observes $Y = DY_1 + (1 - D)Y_0$, and also D and X. For the sake of simplicity, we do not account for uncertainty on potential outcomes. Nevertheless, it would be straightforward to adapt our identification strategy to the case where sectoral decisions depend on expectations of Y_0 and Y_1 rather than on their true values (see D'Haultfœuille and Maurel, 2011). Without loss of generality, we assume that there exists $x_0 \in \text{Supp}(X)$ such that $\psi_0(x_0) = \psi_1(x_0) = 0$.

The generalized Roy models we consider in this section can be used in a broad range of economic settings. The standard Roy model, in which the chosen sector is the one yielding the higher earnings, corresponds to $\eta = 0$ and $G(X) = 0$. This framework also encompasses the Heckman (1974) model of labor market participation. In this latter case, Y_1 corresponds to the logarithm of the potential wage, $G_1(X) = \eta_1 = 0$, $Y_0 = 0$, and $G_0(X)$ (resp. η_0) is the observable (resp. unobservable) part of the logarithm of the reservation wage. More generally, generalized Roy models are well suited for most of the situations in which self-selection between two alternatives is driven by both the relative pecuniary and nonpecuniary returns. They can be used, e.g., to model the decision to attend higher education after graduating from high school, thus extending Willis and Rosen (1979) by accounting for nonpecuniary factors affecting the schooling decision (see, e.g., Carneiro, Hansen, and Heckman, 2003; D'Haultfœuille and Maurel, 2011). Other examples of applications include occupational choice (for the choice between private and public sector, see, e.g., Dagsvik and Strøm, 2006), in addition to migration decisions (see, e.g., Borjas, 1987; Bayer, Khan, and Timmins, 2011) accounting for nonpecuniary factors.[9] Theorem 2.1 can be applied to provide identification of (ψ_0, ψ_1) without exclusion restrictions or large support conditions on the covariates, as the following result shows.

COROLLARY 4.1. *Suppose that* $(\varepsilon_0, \varepsilon_1, \eta) \perp\!\!\!\perp X$, *the suprema of the supports of* ε_0 *and* ε_1 *are infinite, and there exists* $\beta_0, \beta_1 > 0$ *such that* $\mathbb{E}[\exp(\beta_i \varepsilon_i)] < \infty$ *for* $i \in \{0, 1\}$ *and*

$$\lim_{u \to \infty} P(\varepsilon_i + (1 - 2i)\eta \leq a + u | \varepsilon_{1-i} = u) = l_{1-i} > 0 \tag{4.1}$$

for all $a \in \mathbb{R}$ *and* $i \in \{0, 1\}$. *Then* $\psi_0(\cdot)$ *and* $\psi_1(\cdot)$ *are identified.*

Proof. Because $(\varepsilon_0, \varepsilon_1, \eta) \perp\!\!\!\perp X$, condition (4.1) implies that

$$\lim_{u \to \infty} P(Y_1 \geq Y_0 + G(X) + \eta | X = x, Y_1 = u) = l_1.$$

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

In other words,

$$\lim_{u\to\infty} P(D=1|X=x, Y_1=u) = l_1.$$

Thus, we can apply Theorem 2.1 to (D, DY_1, X), and ψ_1 is identified. The same result holds for ψ_0. ■

To the best of our knowledge, this is the first identification result on the effects of covariates in generalized Roy models without exclusion restrictions. Identification without exclusion restrictions of the competing risk model, which is closely related to the standard Roy model, has already been considered in the literature by Heckman and Honore (1989),[10] Abbring and van den Berg (2003), Lee (2006), and Lee and Lewbel (2011). Interestingly, and similarly to our approach, none of these papers requires a large support assumption on the covariates to identify the effect of the covariates. However, all of the strategies proposed in these papers break down when turning to generalized Roy models. Indeed, they rely extensively on the fact that the observed duration is the minimum of potential durations, whereas the observed outcome does not satisfy such a simple property in generalized Roy models.

Identification of (ψ_0, ψ_1) is obtained in Corollary 4.1 under rather mild restrictions on the unobservables. In particular, condition (4.1) can be understood as a moderate dependence assumption between the unobservables. It is automatically satisfied, e.g., if ε_0, ε_1, and η are independent. It also holds if $(\varepsilon_0, \varepsilon_1, \eta)$ is Gaussian, provided that

$$|\mathrm{Cov}(\varepsilon_i, \varepsilon_{1-i} + (2i-1)\eta)| < V(\varepsilon_i), \qquad i \in \{0,1\}.$$

This condition does not put drastic restrictions on the dependence between the unobservables. For instance, it is satisfied in the standard log-normal Roy model if $V(\varepsilon_0) = V(\varepsilon_1)$, as long as $(\varepsilon_0, \varepsilon_1)$ is nondegenerate. It also holds, e.g., in the Heckman (1974) empirical application to labor market participation of married women, although the estimated correlation between ε_1 and η_0 is quite large.[11]

5. CONCLUDING REMARKS

This paper shows that identification of endogenous sample selection models can be achieved without instruments by letting the outcome, not a covariate, tend to the upper bound of its support. The main condition, apart from the exogeneity of the covariates, is the "independence at the limit" of the selection variable and the covariates. In particular, unlike Chamberlain (1986) and Lewbel (2007), our identification strategy does not rely on the existence of a covariate with a large support. Besides, another attractive feature of the proposed identification strategy lies in its testability. Interestingly, and even if a formal procedure remains to be developed, heuristic investigations suggest that our strategy can be tested with typical sample sizes in economics. Noteworthy also, our identification proof is constructive, and

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

an estimator of $\psi(\cdot)$ and $\sigma(\cdot)$ could be based, e.g., on (2.8). One possible route for estimation would be to use trimmed means, as in Heckman (1990) and Schafgans and Zinde-Walsh (2002). In this case, we conjecture that the rate of convergence would depend on the thickness of the tail of the distribution of the outcome, as in Andrews and Schafgans (1998), Schafgans and Zinde-Walsh (2002), and Khan and Tamer (2010). We leave this interesting issue for future research.

NOTES

1. These restrictions entail that the probability of selection tends to zero or one when the special regressor takes arbitrarily large values.

2. To see why such a normalization is always possible, let $\widetilde{\psi}(x) = \psi(x) - \psi(x_0)\sigma(x)/\sigma(x_0)$, $\widetilde{\sigma}(x) = \sigma(x)/\sigma(x_0)$, and $\widetilde{\varepsilon} = \psi(x_0) + \sigma(x_0)\varepsilon$. Then $Y^* = \widetilde{\psi}(X) + \widetilde{\sigma}(X)\widetilde{\varepsilon}$, with $\widetilde{\psi}(x_0) = 0$ and $\widetilde{\sigma}(x_0) = 1$.

3. Instead of supposing $\mathbb{E}(\exp(\beta\varepsilon)) < \infty$ for some $\beta > 0$, we could impose the slightly weaker condition that the survival function $S_{\exp(\varepsilon)}$ of $\exp(\varepsilon)$ is not slowly varying at infinity. We could also impose the even weaker condition that $\mathbb{E}(\varepsilon^\beta) < \infty$ for some $\beta > 0$, but this would come at the price of imposing a finite lower bound on the distribution of ε.

4. Neither additive separability nor monotonicity in η of the index in (2.2) is needed. If $D = \mathbb{1}\{\varphi(X, \eta) \geq 0\}$, the same reasoning applies provided that for all x, $\liminf_{u\to\infty} \varphi(x, u) > 0$. On the other hand, Assumption 3 fails to hold in general when $h(\cdot)$ is bounded.

5. One might not know a priori whether Assumption 3 holds for small or large values of the outcome. We indicate in Section 3 how to test for this.

6. We thank a referee for pointing us to this analogy.

7. The identification proof relies on a finite mean assumption on the unobserved heterogeneity, in a similar spirit to the tail restrictions that we impose in Assumption 2.

8. To simplify the discussion, we consider here only the case where $\overline{y}_x = \infty$. A result analogous to Theorem 3.1 holds otherwise, using the function r defined in the proof of Theorem 2.1 instead of q.

9. Generalized Roy models are also used as a structural underlying framework for the estimation of treatment effects, with D corresponding in that case to the treatment status and $G + \eta$ to the cost of receiving treatment (see Heckman and Vytlacil, 2005). Here however, we cannot recover average treatment effects in general because we do not identify $\mathbb{E}(\varepsilon_0)$ and $\mathbb{E}(\varepsilon_1)$. Yet, the distribution of treatment effects can be point or set identified under additional restrictions (see D'Haultfœuille and Maurel, 2011).

10. Heckman and Honore (1989) use exclusion restrictions but only to identify the distribution of the underlying durations. Their proof shows that the effects of covariates are identified without such restrictions.

11. Heckman obtains, when considering annual hours worked (see Table 1, p. 687), $V(\varepsilon_1) = 0.452^2$, $V(\eta_0) = 0.532^2$, and $\text{Corr}(\varepsilon_1, \eta_0) = 0.654$, so that $|\text{Cov}(\varepsilon_1, -\eta_0)| = 0.157 < V(\varepsilon_1) = 0.204$ (in the case of the sample selection model, $\varepsilon_0 = \eta_1 = 0$, and one need not verify that $|\text{Cov}(\varepsilon_0, \varepsilon_1 - \eta)| < V(\varepsilon_0)$). The same holds for annual weeks worked (see Heckman's Table 2, p. 687).

REFERENCES

Abbring, J. (2010) Identification of dynamic discrete choice models. *Annual Review of Economics* 2, 367–394.

Abbring, J. & G. van den Berg (2003) The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society, Series B* 65, 701–710.

Andrews, D.K. & M. Schafgans (1998) Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press

Bayer, P.J., S. Khan, & C. Timmins (2011) Nonparametric identification and estimation in a Roy model with common nonpecuniary returns. *Journal of Business & Economic Statistics* 29, 201–215.

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

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

Chamberlain, G. (1986) Asymptotic efficiency in semiparametric model with censoring. *Journal of Econometrics* 32, 189–218.

Dagsvik, J. & S. Strøm (2006) Sectoral labour supply, choice restrictions and functional form. *Journal of Applied Econometrics* 21, 803–826.

D'Haultfœuille, X. & A. Maurel (2011) Inference on an Extended Roy Model, with an Application to Schooling Decisions in France. Working paper, Duke University.

Elbers, C. & G. Ridder (1982) True and spurious duration dependence The identifiability of the proportional hazard model. *Review of Economic Studies* 49, 403–409.

Heckman, J.J. (1974) Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–694.

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

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

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

Khan, S. & E. Tamer (2010) Irregular identification, support conditions and inverse weight estimation. *Econometrica* 78, 2021–2042.

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

Lee, S. & A. Lewbel (2011) Nonparametric Identification of Accelerated Failure Time Competing Risks Models. Working paper, Boston College.

Lewbel, A. (2007) Endogenous selection or treatment model estimation. *Journal of Econometrics* 141, 777–806.

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

Schafgans, M. & V. Zinde-Walsh (2002) On intercept estimation in the sample selection model. *Econometric Theory* 18, 40–50.

Vella, F. (1998) Estimating models with sample selection bias: A survey. *Journal of Human Resources* 33, 127–169.

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

https://doi.org/10.1017/S026646661200028X Published online by Cambridge University Press