"A triangular treatment effect model with random coefficients in the selection equation"

Eric Gautier, Stefan Hoderlein

A TRIANGULAR TREATMENT EFFECT MODEL WITH RANDOM COEFFICIENTS IN THE SELECTION EQUATION

ERIC GAUTIER AND STEFAN HODERLEIN

ABSTRACT. This paper considers treatment effects under endogeneity with complex heterogeneity in the selection equation. We model the outcome of an endogenous treatment as a triangular system, where both the outcome and first-stage equations consist of a random coefficients model. The first-stage specifically allows for nonmonotone selection into treatment. We provide conditions under which marginal distributions of potential outcomes, average and quantile treatment effects, all conditional on first-stage random coefficients, are identified. Under the same conditions, we derive bounds on the (conditional) joint distributions of potential outcomes and gains from treatment, and provide additional conditions for their point identification. All conditional quantities yield unconditional effects (*e.g.*, the average treatment effect) by weighted integration.

1. INTRODUCTION

To evaluate the effect of a treatment, it is common in social sciences to rely on non-experimental data. In such a setup, ignoring self selection into treatment results in a misleading assessment of the effectiveness of the treatment, as causal effects may be confounded with the effect of endogenous selection. Another important feature of such real-world applications is the heterogeneity of the effect of treatment across individuals. The most prominent identification strategy relies on the (instrument) *monotonicity* assumption of Imbens and Angrist (1994). It means that there are *no defiers* in the sense

that as an instrument is shifted from z to z' (say, a voucher to participate in treatment is randomly allocated), individuals either stay with their treatment choice or move into treatment, but no individual reverses his decision. Vytlacil (2002) shows that this assumption is equivalent to modeling the selection as an additively separable latent index model with a single unobservable, where individuals are increasingly likely to be in the treatment group as the value of this unobservable increases.

The monotonicity assumption is often too restrictive. The incentives individuals face might be complex and commonly cost-benefit trade-offs, as exemplified in the Roy model, are being performed. These considerations depend on several parameters, all of which may be expected to vary across the population. As such, more complex models of endogenous selection that explicitly allow for complex unobserved heterogeneity should be considered. A particularly natural specification for the selection equation that accomplishes this goal is given by a random coefficients selection equation. Heckman and Vytlacil (2005) (HV henceforth) call it the benchmark nonseparable, nonmonotic model of treatment choice and emphasizes the importance of being equally flexible in terms of unobserved heterogeneity in the outcome and selection (or first-stage) equations. In this paper, we consider a triangular system where the outcome equation is a linear random coefficients model with a binary endogenous regressor (*i.e.*, the treatment status) and the selection equation is a nonlinear random coefficients binary choice model. It allows for a population with both compliers and defiers. Related literature includes Beran and Millar (1994), Beran, Feuerverger and Hall (1996) and Hoderlein, Klemelä and Mammen (2010) for the linear random coefficients model, Lewbel and Pendakur (2012) for a different nonlinear specification, Ichimura and Thomson (1998) and Gautier and Kitamura (2013) for the random coefficients binary choice model.

Under the monotonicity assumption, the Marginal Treatment Effect (MTE, Björklund and Moffitt (1987) and HV (2005)) is a key structural parameter to recover a large variety of effects. In our framework without the monotonicity assumption, we show that marginal distributions of the potential outcomes, a generalization of the MTE and of the Quantile

Treatment Effect (QTE, Abadie, Angrist and Imbens (2002)), conditional on the first-stage random coefficients, are identified. Under the same assumptions, we obtain bounds on the joint distribution of potential outcomes and the distribution and variance of the gains, conditional on the first-stage random coefficients. We also provide conditions for their identification. The conditional effects that we obtain are structural parameters of interest in their own right, as we argue below. They are also building blocks for average, quantile or distributional effects.

2. THE THEORETICAL SETUP

Throughout this paper, we make use of the following notations. The conditional distribution of a random vector A given $B = b$ is denoted by $\mathbb{P}_{A|B}(\cdot|b)$; its conditional CDF by $F_{A|B}(\cdot|b)$; its conditional PDF by $f_{A|B}(\cdot|b)$ when it exists; its conditional expectation (resp. variance) by $\mathbb{E}[A|B = b]$ (resp. $Var(A|B = b)$) when A is scalar; and the support of the conditional distribution by $\text{supp}\left(\mathbb{P}_{A|B}(\cdot|b)\right)$. The unconditional quantities are denoted without $|B$. We use the notation $A \perp B|X$ when A and B are independent given X, a.s. for almost surely, a.e. for almost everywhere, $\sigma(A)$ for the sigma algebra generated by A. All random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and ω is an element of Ω. We also denote by $\mathbb{1}$ the indicator function, by $\text{Int}(A)$ the interior of a set A, by $\| \cdot \|$ the Euclidian norm, and by A_{2-L} the subvector $(A_2, \ldots, A_L)$ of $(A_1, \ldots, A_L)$.

In the treatment effects framework, Y_0 is the outcome in the control group, Y_1 is the outcome in the treated group, $\Delta = Y_1 - Y_0$ is the gain from treatment or treatment effect, D is the binary variable which is equal to 1 when treatment has been selected and 0 otherwise, and Z is a vector of instruments. We denote by $D_z(\omega)$ the value of $D(\omega)$ if $Z(\omega) = (Z_1(\omega), \ldots, Z_L(\omega))$ were externally set to z (one has $D(\omega) = D_{Z(\omega)}(\omega)$).

2.1. **Relaxing monotonicity.** Endogenous selection into treatment is often dealt with by specifying a selection equation with a single unobservable. In its general form, it can be written as $D = \mathbb{1}\{\mu(F(Z), \Theta) > 0\}$, where Z is independent of (Y_0, Δ, Θ), F is measurable, $F(Z)$ and Θ are scalar, and μ is increasing in $F(Z)$. This equation can be

written as an additively separable latent index model which, as shown in Vytlacil (2002), is equivalent to the LATE assumptions of Imbens and Angrist (1994):

(L1) $\forall z \in \text{supp}(\mathbb{P}_Z)$ $Z \perp (Y_0, Y_1, D_z)$ and $z \to \mathbb{P}(D = 1 | Z = z)$ is not constant;

(L2) $\forall z, z' \in \text{supp}(\mathbb{P}_Z)$, $\forall \omega \in \Omega$ $D_z(\omega) \le D_{z'}(\omega)$ or $\forall \omega \in \Omega$ $D_{z'}(\omega) \le D_z(\omega)$.

(L2) is called monotonicity. A few papers recently aimed at relaxing (L2). Small and Tan (2007) gives a formula for the Local Average Treatment Effect (LATE) as a function of quantities conditional on a latent variable U, but does not study their identification, and gives a degenerate example where $U = (Y_1, Y_0)$. Klein (2010) considers a selection equation with two unobservables and the order of the bias of classical estimators when one unobservable goes to zero. Fox and Gandhi (2011) studies identification of the distribution of unobserved heterogeneity in Roy models with discretely supported random coefficients, without a random intercept in the selection equation.

2.2. **The model.** In this paper, we consider the triangular random coefficients model

$$(2.1) \qquad Y = Y_0 + \Delta D,$$

$$(2.2) \qquad D = \mathbb{1}\left\{ \Theta - \Gamma_1(Z_1 + g(Z_2, \ldots, Z_L)) - \sum_{l=2}^{L} \Gamma_l f_l(Z_l) > 0 \right\}, \quad L \ge 2$$

where $f_2, \ldots, f_L$ and g are unknown. The random elements Y, D, Z are observed, while Y_0, Δ, Θ and Γ are unobserved. The unobservables can be arbitrarily dependent. We do not assume the existence of densities or rely on parametric assumptions.

When available, conditioning variables X, with common values when D is exogenously set to 0 or 1, can be introduced for two reasons: (1) to obtain effects for population subgroups (in which case they are not important in our identification arguments); (2) to justify the exogeneity of the instruments, or conditional independence assumptions.

The outcome equation (2.1) is a linear random coefficients equation with random coefficients (Y_0, Δ) and binary endogenous regressor D, namely D and (Y_0, Δ) are dependent. This specification imposes no restriction on the outcome equation. The structural parameter in this model is $F_{Y_0, \Delta}$. It is in a one-to-one relationship with the joint distribution of

potential outcomes F_{Y_0,Y_1}. Important functionals of $F_{Y_0,\Delta}$ include: the average marginal effect $\mathbb{E}[\Delta]$, also called average treatment effect (ATE), the variance of treatment effects, and the proportion of people who benefit from treatment $\mathbb{P}(\Delta > 0)$, among others.

Example RCR: A generalized Roy model with uncertainty and random coefficients. Individuals select treatment if their *ex-ante* gain from treatment $\mathbb{E}[\Delta|\mathcal{I}]$ exceeds cost. The *ex-post* gain from treatment is Δ. The sigma-field $\mathcal{I}$ is the information set individuals have at their disposal at the time of their decision to participate in the treatment. Their cost is defined as $\Gamma_0 + \Gamma_1(Z_1 + g(Z_{2-L})) + \sum_{l=2}^{L} \Gamma_l f_l(Z_l)$. The instruments are cost shifters and the random coefficients reflect the heterogeneous impact of nonlinear transformations on cost. We assume, for simplicity, that the individuals know the value of their random coefficients. We obtain model (2.2) where $\Theta = \mathbb{E}[\Delta|\mathcal{I}] - \Gamma_0$. The *ex-ante* returns from treatment is EAR $:= \mathbb{E}[\Delta|\mathcal{I}] - \Gamma_0 - \Gamma_1(Z_1 + g(Z_{2-L})) - \sum_{l=2}^{L} \Gamma_l f_l(Z_l)$, while the *ex-post* returns is EPR $:= \Delta - \Gamma_0 - \Gamma_1(Z_1 + g(Z_{2-L})) - \sum_{l=2}^{L} \Gamma_l f_l(Z_l)$ $\qquad \square$

Model (2.1)-(2.2) nests the benchmark model put forward in HV (2005) and allows more general treatment choices than (L2) when Γ_{2-L} has one nondegenerate coordinate.

Because the individual values of a vector of random coefficients cannot be obtained when its dimension exceeds one, we cannot rely on a type of control function approach (see Newey, Powell and Vella (1999) and Imbens and Newey (2009)). Kasy (2011) shows that, when the endogenous regressor is continuous, the control function approach cannot be used with a random coefficients first-stage. Identification of causal effect requires new tools, which are the tools that we now develop.

2.3. **Main Identifying Assumptions.** In this section, we present our main set of identifying assumptions. Throughout this article, we make the following sign restriction.

Assumption 2.1. $Z \perp (\Theta, \Gamma)|X$ and Γ_1 has a sign and is not 0 a.s.

Conditioning on $Z = z$ and $X = x$ for $(z, x) \in \text{Int}(\text{supp}(\mathbb{P}_{Z,X}))$, increasing z_1 increases (resp. decreases) $\mathbb{P}(D = 1|Z = z, X = x)$ if Γ_1 is negative (resp. positive). Therefore, the sign of Γ_1 can be identified. Since it is possible to change Z_1 in $-Z_1$, we now assume

that $\Gamma_1 > 0$ a.s. We can then divide both sides of the inequality in (2.2) by Γ_1 and work with the coefficients $\overline{\Theta} := \Theta/\Gamma_1$, $\overline{\Gamma}_2 := \Gamma_2/\Gamma_1, \ldots, \overline{\Gamma}_L := \Gamma_L/\Gamma_1$.

The same argument shows that we do not have to know which instrument plays the role of Z_1. Indeed, it can be solely determined by $\mathbb{P}(D = 1 | Z = z, X = x)$. When D is college attendance, tuition or distance to university could play the role of Z_1.

In the linear random coefficients binary choice model, assuming full support of the instruments and without invoking variables X, a sufficient condition for identification of the distribution of the random coefficients $(\Theta, \Gamma_1^0, \ldots, \Gamma_L^0)$ up to scale is that, for a unit vector s, $s^T(\Theta, \Gamma_1^0, \ldots, \Gamma_L^0) > 0$ a.s. This means that that there exists a value z of the instruments such that $D_z = 1$ a.s. (in which case we take $s = (1, z)/\|(1, z)\|$) or $D_z = -1$ a.s. (in which case we take $s = -(1, z)/\|(1, z)\|$). This condition is satisfied if one coefficient has a sign. More generally, it means that by working with the vector of transformed instrument AZ where A is invertible and such that the first row of $(A^T)^{-1}$ is s^T, we obtain the new random coefficients Θ and $\Gamma = (A^T)^{-1}\Gamma^0$ where $\Gamma_1 > 0$ a.s. In equation (2.2), we replace the linear transformation by a nonlinear transformation, where Z_1 is transformed into $Z_1 + g(Z_{2-L})$.

We now invoke moments, instrument independence, and large support assumptions.

Assumption 2.2. (A1) $\mathbb{E}[|Y_0| + |\Delta|] < \infty$;
(A2) $Z \perp (Y_d, \overline{\Theta}, \overline{\Gamma}_{2-L}) | X$ for $d = 0, 1$;
(A3) $0 < \mathbb{P}(D = 1 | X) < 1 \quad a.s.$;
(A4) $\forall (z_{2-L}, x) \in \text{supp}\left(\mathbb{P}_{Z_{2-L}, X}\right)$,

$$\text{supp}\left(\mathbb{P}_{Z_1 | Z_{2-L}, X}(\cdot | z_{2-L}, x)\right) \supseteq \text{supp}\left(\mathbb{P}_{\overline{\Theta} - g(z_{2-L}) - \sum_{l=2}^{L} \overline{\Gamma}_l f_l(z_l) | X}(\cdot | x)\right).$$

(A1) allows us to consider conditional expectations of Y_0 and Δ. (A2) allows $(Y_0, \overline{\Theta}, \overline{\Gamma}_{2-L})$ and $(Y_1, \overline{\Theta}, \overline{\Gamma}_{2-L})$ to depend on Z, as long as we have at hand control variables X which yield independence. These variables X can either be observed directly, or they may be constructed from another additional equation, say, as control functions. Moreover, they need not exist at all. (A3) states that there is a fraction of the population that participates in treatment, and one that does not. (A4) means that Z_1 has a large enough

support. A similar assumption is made in Lewbel (2007). (A4) implies the existence of unselected samples when Z_1 (only) approaches the lower and upper bounds of its support, which could be tested. Traditional "identification at infinity" makes inefficient use of the data, while (A4) allows to build estimators that use the entire data (see Gautier and Hoderlein (2012)). Finally, though "identification at infinity" arguments allow to obtain effects that depend on F_{Y_d} for $d = 0, 1$, they do not yield the structural parameters conditional on first-stage unobservables.

In a preliminary step, these assumptions may be taken to identify the nonlinear elements in (2.2). Observe that using (A2), for every $(z, x) \in \text{supp}(\mathbb{P}_{Z,X})$,

$$\mathbb{P}(D = 0 | Z = z, X = x) = \mathbb{P}\left(\overline{\Theta} - g(z_{2-L}) - \sum_{l=2}^{L} \overline{\Gamma}_l f_l(z_l) \le z_1 | X = x\right).$$

Therefore, using (A4), identification in (2.2) is implied by identification in

$$W = g(Z_{2-L}) + \sum_{l=2}^{L} \overline{\Gamma}_l f_l(Z_l) - \overline{\Theta}$$

where $Z_{2-L} \perp (\overline{\Gamma}_{2-L}, \overline{\Theta}) | X$. We now make integrability and location normalizations.

Assumption 2.3. (N1) $\mathbb{E}\left[|g(Z_{2-L})| + \sum_{l=2}^{L} |\overline{\Gamma}_l| + |f_l(Z_l)|\right] < \infty$;
(N2) For $l = 2, \ldots, L$, $\mathbb{E}[g(Z_{2-L})|Z_l, X] = 0$ a.s.;
(N3) For $l = 2, \ldots, L$, $\mathbb{E}[\overline{\Gamma}_l|X] = 1$ a.s.;
(N4) For $l = 2, \ldots, L$, $\mathbb{E}[f_l(Z_l)|X] = 0$ a.s.

The following theorem shows identification of the nonlinear elements in (2.2).

Theorem 2.1. f_l for $l = 2, \ldots, L$ and g are identified under Assumption 2.3.

Writing $\overline{Z}_1 := -Z_1 + g(Z_{2-L})$ and $\overline{Z}_l := f_l(Z_l)$ for $l = 2, \ldots, L$, we now work with

(2.3) $$D = \mathbb{1}\left\{-\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l \overline{Z}_l < \overline{Z}_1\right\}$$

where $\mathbb{E}[\overline{\Gamma}_l|X] = 1$ for $l = 2, \ldots, L$.

When the large support assumption (A4) is not satisfied, it is replaced by (A4'):

(A4'a) Model (2.2) does not have unknown functions;

(A4'b) $\forall x \in \operatorname{supp}(\mathbb{P}_X)$, $\forall R \geq 0$, $\mathbb{E}\left[\exp\left(R|\overline{\Theta}|\right)\middle| X = x\right] < \infty$;

(A4'c) $\forall (z_{2-L}, x) \in \operatorname{supp}\left(\mathbb{P}_{\overline{Z}_{2-L}, X}\right)$, $\operatorname{Int}\left(\operatorname{supp}\left(\mathbb{P}_{Z_1|\overline{Z}_{2-L}, X}(\cdot|z_{2-L}, x)\right)\right)$ is not empty.

Finally, we invoke either of the two following assumptions on $\operatorname{supp}(\mathbb{P}_{\overline{Z}_{2-L}|X})$.

Assumption 2.4. $\forall x \in \operatorname{supp}(\mathbb{P}_X)$, $\operatorname{supp}\left(\mathbb{P}_{\overline{Z}_{2-L}|X}(\cdot|x)\right) = \mathbb{R}^{L-1}$.

Assumption 2.5. $\forall x \in \operatorname{supp}(\mathbb{P}_X)$,

(S1) $\forall R \geq 0$, $\mathbb{E}\left[\exp\left(R\left\|\overline{\Gamma}_{2-L}\right\|\right)\middle| X = x\right] < \infty$;

(S2) $\operatorname{Int}\left(\operatorname{supp}\left(\mathbb{P}_{\overline{Z}_{2-L}|X}(\cdot|x)\right)\right)$ is not empty.

(S1) corresponds to the existence of a moment generating function of $\overline{\Gamma}_{2-L}$, which can be relaxed as in Gaillac and Gautier (2015). Beran and Millar (1994) assumes that the support of the random coefficients, including the random intercept, is compact.

3. PARAMETERS BASED ON THE MARGINALS OF POTENTIAL OUTCOMES

3.1. **Identification of Marginals Conditional on First-stage Unobservables.** The following result is central in our analysis.

Theorem 3.1. Under Assumption 2.2 with (A4) or (A4') and either Assumption 2.4 or 2.5, $F_{Y_d, \overline{\Theta}, \overline{\Gamma}_{2-L}, X}$ for $d = 0, 1$ are identified.

This result shows identification of $F_{Y_0 + \Delta d|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}$ and $F_{Y_0 + \Delta d|X}$ for $d = 0, 1$, and of the average and quantile structural functions. This means that we have resolved the endogeneity issue in model (2.1) entirely and the difficulty in identifying the structural parameter $F_{Y_0, \Delta|X}$ now stems solely from the fact that the regressor is binary.

As will be obvious, Theorem 3.1 may be used to establish identification of many parameters. For instance, it directly yields identification of the following parameter.

Definition 3.1. The Unobservables Conditioned Quantile Treatment Effect (UCQTE) is defined, for $(\theta, \gamma_{2-L}, x)$ in $\operatorname{supp}(\mathbb{P}_{\overline{\Theta}, \overline{\Gamma}_{2-L}, X})$, as

$$\operatorname{UCQTE}(\theta, \gamma_{2-L}, x, \tau) := F_{Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(\cdot|\theta, \gamma_{2-L}, x)^{-1}(\tau) - F_{Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(\cdot|\theta, \gamma_{2-L}, x)^{-1}(\tau).$$

Integrating out the first-stage unobserved heterogeneity, we obtain

$$\text{QTE}(x, \tau) = F_{Y_1|X}(\cdot|x)^{-1}(\tau) - F_{Y_0|X}(\cdot|x)^{-1}(\tau), \ \tau \in (0, 1)$$

without monotonicity or rank invariance (Chernozhukov and Hansen (2005)).

3.2. **The Unobservables Conditioned Average Treatment Effect (UCATE).**

3.2.1. *Definition and Properties.* MTE is a key structural parameter (see HV (2005)) which relies on the additively separable latent index model for the first-stage, thus on monotonicity. To extend this concept to our framework, we introduce the following generalization of the MTE. It is clearly identified from Theorem 3.1.

Definition 3.2. The Unobservables Conditioned Average Treatment Effect (UCATE) is defined, for $(\theta, \gamma_{2-L}, x)$ in $\text{supp}(\mathbb{P}_{\overline{\Theta}, \overline{\Gamma}_{2-L}, X})$, as

$$\text{UCATE}(\theta, \gamma_{2-L}, x) := \mathbb{E}[\Delta|\overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x] \ .$$

UCATE shares the same interpretation and attractive properties as MTE. It is the average effect for a subpopulation with $X = x$ which would be indifferent between participation and nonparticipation in the treatment, if it were exogenously assigned a value z of $\overline{Z}$ such that $-\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l z_l = z_1$. Due to (A2), this parameter is policy invariant (*i.e.*, it is independent of the values of the instruments)

$$\mathbb{E}\left[\Delta|\overline{\Theta}, \overline{\Gamma}_{2-L}, X, \overline{Z}\right] = \mathbb{E}[\Delta|\overline{\Theta}, \overline{\Gamma}_{2-L}, X] \ .$$

Like MTE, UCATE is an economically important parameter. To fix ideas, think of the treatment as attending college, and of the j-th cost factor as college tuition. If an individual has a high value of $\overline{\Gamma}_l$, she is responsive to a change in college tuition. This may be because she is liquidity constrained. For a policy maker who is deciding about tuition rates, the difference in treatment effects across these groups is a key variable in assessing the effect of a change in the incentive structure. For targeted policy measures, it is interesting to see how the effect of treatment varies with unobserved sensitivity to tuition and observed X (*e.g.*, race).

Finally, note that, if $\mathrm{UCATE}(\theta, \gamma_{2-L}, x)$ is not constant in one of the γ_l's, it is an indication that monotonicity does not hold. This allows to test monotonicity.

3.2.2. *Example RCR: UCATE as the Ex-ante Gains From Treatment.* In this model, UCATE is a core structural effect. We invoke the following assumption.

Assumption 3.1. (U1) $\mathcal{I} \supseteq \sigma(\overline{\Gamma}_{2-L}, X)$
(U2) $\sigma(\overline{\Gamma}_{2-L}, X) = \sigma(\Gamma_0, \Gamma_1, \ldots, \Gamma_L, X)$.

Under Assumption 3.1, individuals have perfect knowledge of their cost function, but can be uncertain about the potential outcomes.

(U2) is satisfied when $(\Gamma_0, \Gamma_1, \ldots, \Gamma_L) = \varphi(\overline{\Gamma}_{2-L})$ for some measurable function φ. This restricts the unobserved heterogeneity entering (2.2). However, it does not restrict its dimension. Example RCR also involves $\mathbb{E}[\Delta|\mathcal{I}]$ in (2.2), so that, when the superset symbol in (U1) is strict, (2.2) depends on a nondegenerate vector of unobservables of dimension L. A sufficient condition for (U2) to hold is that one Γ_l for $l = 1, \ldots, L$ is $\sigma(X)$-measurable and $\mathbb{P}(\Gamma_l = 0) = 0$, and Γ_0 is $\sigma(\Gamma_1, \ldots, \Gamma_L, X)$-measurable. Still, it is a richer specification than the extended Roy model where the cost function is not random.

Define the errors η_d for $d = 0, 1$ as

$$(3.1) \qquad \eta_d := \mathbb{E}[Y_d|\mathcal{I}] - \mathbb{E}[Y_d|\overline{\Gamma}_{2-L}, X] .$$

When Example RCR models sectorial choice, they could be interpreted as sector specific skills known by the individuals at the time the choice is made and unobserved by the econometrician. Because $\overline{\Theta} = (\mathbb{E}[Y_1 - Y_0|\mathcal{I}] - \Gamma_0)/\Gamma_1 = (\mathbb{E}[Y_1 - Y_0|\overline{\Gamma}_{2-L}, X] + \eta_1 - \eta_0 - \Gamma_0)/\Gamma_1$, Assumption 3.1 implies

$$(3.2) \qquad \sigma(\overline{\Theta}, \overline{\Gamma}_{2-L}, X) = \sigma(\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X) \subseteq \sigma(\overline{\Gamma}_{2-L}, \eta_1, \eta_0, X) \subseteq \mathcal{I} .$$

Even if $\sigma(\overline{\Theta}, \overline{\Gamma}_{2-L}, X)$ is a strict subset of $\mathcal{I}$, the following result holds.

Theorem 3.2. Under Assumption 3.1 with (A4) or (A4'), $\mathrm{UCATE}(\overline{\Theta}, \overline{\Gamma}_{2-L}, X) = \mathbb{E}[\Delta|\mathcal{I}]$. Moreover, if $\mathbb{E}[\Delta|\mathcal{I}]$ is not $\sigma(\overline{\Gamma}_{2-L}, X)$-measurable, $F_{\mathrm{EAR}|Z,X}$ is identified.

3.2.3. *UCATE as a Building Block to Obtain Treatment Effects that Depend on Averages.* Like MTE, UCATE is a building block to obtain treatment effect parameters that depend on averages. For example, the ATE and Treatment on the Treated (TT) are given by

$$\text{ATE}(x) = \int_{\mathbb{R}^L} \text{UCATE}(\theta, \gamma_{2-L}, x) d\mathbb{P}_{\overline{\Theta}, \overline{\Gamma}_{2-L}|X}(\theta, \gamma_{2-L}|x) \ ,$$

$$\text{TT}(x) = \int_{\mathbb{R}^L} h_{\text{TT}}(\theta, \gamma_{2-L}, x) \text{UCATE}(\theta, \gamma_{2-L}, x) d\mathbb{P}_{\overline{\Theta}, \overline{\Gamma}_{2-L}|X}(\theta, \gamma_{2-L}|x) \ ,$$

where $h_{\text{TT}}(\theta, \gamma_{2-L}, x) = \mathbb{E}\left[D|X = x\right]^{-1} \mathbb{E}\left[\left. \mathbb{1}\left\{-\theta + \sum_{l=2}^{L} \gamma_l \overline{Z}_l < \overline{Z}_1\right\} \right| X = x\right]$.

UCATE also allows to derive Policy Relevant Treatment Effect (PRTE) parameters. In our setup where an individual self-select into treatment, these parameters inform the decision maker about the effect of a change in the structure of the variables or incentives she can control, namely the distribution of instruments. For example, the alternative policy can consist in changing Z_1 to $Z_1 + \alpha$ (*e.g.*, increasing tuition by α). HV (2001) formalizes this idea. Denote by $\mathbb{E}_\alpha$ the expectation under a policy α and $\mathbb{E}_0$ the expectation under the baseline policy. The PRTE parameter is defined, for x in $\text{supp}(\mathbb{P}_X)$, as

$$\text{PRTE}(\alpha, x) := \frac{\mathbb{E}_\alpha[Y|X = x] - \mathbb{E}_0[Y|X = x]}{\mathbb{E}_\alpha[D|X = x] - \mathbb{E}_0[D|X = x]} \ .$$

Marginal Policy Relevant Treatment Effects (MPRTE, Carneiro, Heckman and Vytlacil (2010)) are defined as

$$\text{MPRTE}(x) := \lim_{\alpha \to 0} \text{PRTE}(\alpha, x) \ .$$

In our setup, $\text{PRTE}(\alpha, x)$ can be obtained from UCATE like $\text{TT}(x)$ replacing the weight $h_{\text{TT}}(\theta, \gamma_{2-L}, x)$ by $h_{\text{PRTE}}(\theta, \gamma_{2-L}, \alpha, x)$ defined by

$$\frac{\mathbb{E}_\alpha\left[\left.\mathbb{1}\left\{-\theta + \sum_{l=2}^{L} \gamma_l \overline{Z}_l < \overline{Z}_1\right\} \right| X = x\right] - \mathbb{E}_0\left[\left.\mathbb{1}\left\{-\theta + \sum_{l=2}^{L} \gamma_l \overline{Z}_l < \overline{Z}_1\right\} \right| X = x\right]}{\mathbb{E}_\alpha\left[\left.\mathbb{1}\left\{-\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l \overline{Z}_l < \overline{Z}_1\right\} \right| X = x\right] - \mathbb{E}_0\left[\left.\mathbb{1}\left\{-\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l \overline{Z}_l < \overline{Z}_1\right\} \right| X = x\right]} \ .$$

In a similar manner, $\mathrm{MPRTE}(x)$ can be obtained using the weight

$$h_{\mathrm{MPRTE}}(\theta, \gamma_{2-L}, x) = \frac{\partial_\alpha \, \mathbb{E}_\alpha \left[\mathbb{1}\left\{ -\theta + \sum_{l=2}^{L} \gamma_l \overline{Z}_l < \overline{Z}_1 \right\} \Big| X = x \right]\Big|_{\alpha=0}}{\partial_\alpha \, \mathbb{E}_\alpha \left[\mathbb{1}\left\{ -\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l \overline{Z}_l < \overline{Z}_1 \right\} \Big| X = x \right]\Big|_{\alpha=0}} \; .$$

The weights h_{TT}, h_{PRTE} and h_{MPRTE} can be obtained by Monte-Carlo techniques.

4. Parameters Based on the Joint Distribution of Potential Outcomes

In this section, we consider identification of the structural parameter $F_{Y_0, \Delta|X}$ in model (2.2). The importance of distributional treatment effects is emphasized by Heckman, Smith and Clements (1997) (HSC henceforth) and Abbring and Heckman (2007). We present partial identification results under the same assumptions as in Section 3 and present assumptions that allow to identify distributional effects.

4.1. **Partial Identification of Distributional Effects.**
Theorem 3.1 yields that, under our maintained assumptions, $F_{Y_d|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}$ for $d = 0, 1$ are identified. Under the same assumptions, we obtain sharp bounds on structural parameters that depend on $F_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}$. We use the following notations:

$$F_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}^L(y_0, y_1|\theta, \gamma_{2-L}, x) := \max\left\{ F_{Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0|\theta, \gamma_{2-L}, x) + F_{Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_1|\theta, \gamma_{2-L}, x) - 1, 0 \right\}$$

$$F_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}^U(y_0, y_1|\theta, \gamma_{2-L}, x) := \min\left\{ F_{Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0|\theta, \gamma_{2-L}, x), F_{Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_1|\theta, \gamma_{2-L}, x) \right\}$$

$$F_{\Delta|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}^L(\delta|\theta, \gamma_{2-L}, x) := \sup_{y \in \mathbb{R}} \max\left\{ F_{Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y|\theta, \gamma_{2-L}, x) - F_{Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y - \delta|\theta, \gamma_{2-L}, x), 0 \right\}$$

$$F_{\Delta|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}^U(\delta|\theta, \gamma_{2-L}, x) := 1 + \inf_{y \in \mathbb{R}} \min\left\{ F_{Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y|\theta, \gamma_{2-L}, x) - F_{Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y - \delta|\theta, \gamma_{2-L}, x), 0 \right\}$$

and $F_{Y_0, Y_1}^L(y_0, y_1|x)$, $F_{Y_0, Y_1}^U(y_0, y_1|x)$, $F_\Delta^L(\delta|x)$, and $F_\Delta^U(\delta|x)$ are the expectation of above quantities evaluated at the random vectors. For example, we define

$$F_{Y_0, Y_1}^L(y_0, y_1|x) := \int_{\mathbb{R}^L} F_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}^L(y_0, y_1|\theta, \gamma_{2-L}, x) d\mathbb{P}_{\overline{\Theta}, \overline{\Gamma}_{2-L}|X}(\theta, \gamma_{2-L}|x) \; .$$

Theorem 4.1. Under the assumptions of Theorem 3.1, for $(\theta, \gamma_{2-L}, x) \in \mathrm{supp}(\mathbb{P}_{\overline{\Theta}, \overline{\Gamma}_{2-L}, X})$, $(y_0, y_1, \delta) \in \mathbb{R}^3$,

$$F^L_{Y_0, Y_1 | \overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0, y_1 | \theta, \gamma_{2-L}, x) \leq F_{Y_0, Y_1 | \overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0, y_1 | \theta, \gamma_{2-L}, x) \leq F^U_{Y_0, Y_1 | \overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0, y_1 | \theta, \gamma_{2-L}, x)$$

$$F^L_{\Delta | \overline{\Theta}, \overline{\Gamma}_{2-L}, X}(\delta | \theta, \gamma_{2-L}, x) \leq F_{\Delta | \overline{\Theta}, \overline{\Gamma}_{1-2}, X}(\delta | \theta, \gamma_{2-L}, x) \leq F^U_{\Delta | \overline{\Theta}, \overline{\Gamma}_{2-L}, X}(\delta | \theta, \gamma_{2-L}, x) \ .$$

All bounds are sharp. Bounds on functionals of the unobservables conditioned joint distribution of potential outcomes can be obtained like in Fan, Guerre and Zhu (2014).

Corollary 4.1. Under the assumptions of Theorem 3.1, for $x \in \mathrm{supp}(\mathbb{P}_X)$, $(y_0, y_1, \delta) \in \mathbb{R}^3$,

$$F^L_{Y_0, Y_1 | X}(y_0, y_1 | x) \leq F_{Y_0, Y_1 | X}(y_0, y_1 | x) \leq F^U_{Y_0, Y_1 | X}(y_0, y_1 | x)$$

$$F^L_{\Delta | X}(\delta | x) \leq F_{\Delta | X}(\delta | x) \leq F^U_{\Delta | X}(\delta | x) \ .$$

Bounds on the above quantities have been obtained in the case of randomized experiments or selection on observables (see HSC (1997) for F_{Y_0, Y_1}, Fan and Park (2010) and Firpo and Ridder (2008) for F_Δ). The bounds of this section hold when there is endogenous selection into treatment and without mononoticity. It is possible to deduce bounds which are unconditional on (some) X by integration against $\mathbb{P}_X$. These are sharper than the ones obtained without conditioning (see Firpo and Ridder (2008)). Similarly, the bounds of Corollary 4.1 are sharper than bounds obtained without monotonicity (see Kitagawa (2009) for F_{Y_0, Y_1}) and without conditioning on the first-stage unobservables.

4.2. **Point Identification of Distributional Effects.** We start by strengthening (A2):

$$(\text{A2'}) \quad Z \perp (Y_0, Y_1, \overline{\Theta}, \overline{\Gamma}_{2-L}) \,| X.$$

We now present three alternative identifying assumptions and give sufficient conditions for them to hold in the case of Example RCR under Assumption 3.1.

Assumption 4.1. $Y_0 \perp Y_1 \,| \overline{\Gamma}_{2-L}, X$.

Assumption 4.2. $Y_0 \perp Y_1 \,| \overline{\Theta}, \overline{\Gamma}_{2-L}, X$.

Assumption 4.3. (i) $Y_0 \perp \Delta \,| \overline{\Theta}, \overline{\Gamma}_{2-L}, X$;

(ii) Either (1) for a.e. $t \in \mathbb{R}$, $\mathbb{E}[e^{itY_0}|\overline{\Theta}, \overline{\Gamma}_{2-L}, X] \neq 0$ a.s.
   or (2) $\forall R \geq 0$, $\mathbb{E}\left[\exp\left(R|\Delta|\right)|\overline{\Theta}, \overline{\Gamma}_{2-L}, X\right] < \infty$ a.s.

HSC (1997) makes a similar assumption as Assumption 4.3 (i), but it is not conditional on the first-stage unobservables. Condition (ii) (1) is the condition in Devroye (1989) for the deconvolution problem $Y_1 = \Delta + Y_0$ where Y_0 is the error with known distribution. Condition (ii) (2) imposes no restriction on the distribution of Y_0.

We can relabel state 0 as state 1 in the first two assumptions, as well as in Assumption 4.3 (i) if the agents have perfect foresight of the outcomes in both states. This is important when the two states are two sectors in the economy.

Consider now Example RCR. Define $\epsilon_d := Y_d - \mathbb{E}[Y_d|\mathcal{I}]$ for $d = 0, 1$, which could be interpreted as state specific unexpected shocks. Consider the three following assumptions:

(B1) $\eta_0 + \epsilon_0 \perp \eta_1 + \epsilon_1 \ |\overline{\Gamma}_{2-L}, X$;

(B2) Assumption 3.1 holds, $\epsilon_0 \perp \epsilon_1 \ |\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X$, and η_0 is $\sigma(\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X)$ measurable;

(B3) Assumption 3.1 holds and $\eta_0 + \epsilon_0 \perp \epsilon_1 - \epsilon_0 \ |\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X$.

(B1) and (B2) restrict the unobserved heterogeneity in (2.2). (B1) implies

$$Y_0 - \mathbb{E}[Y_0|\overline{\Gamma}_{2-L}, X] \perp Y_1 - \mathbb{E}[Y_1|\overline{\Gamma}_{2-L}, X] \ |\overline{\Gamma}_{2-L}, X \ ,$$

and thus Assumption 4.1. (B2) implies

$$Y_0 - \mathbb{E}[Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X] \perp Y_1 - \mathbb{E}[Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X] \ |\overline{\Theta}, \overline{\Gamma}_{2-L}, X$$

and thus Assumption 4.2. (B3) implies Assumption 4.3 (i). It is the less restrictive on the unobserved heterogeneity entering the first-stage equation. It is satisfied, for example, when there is no uncertainty on the outcome in the base state (i.e., $\epsilon_0 = 0$) and $\epsilon_1 \perp \eta_0, \eta_1, \overline{\Gamma}_{2-L}|X$. Assumptions 4.1 - 4.3 (i) can hold more generally, in particular without Assumption 3.1 which restricts the unobserved heterogeneity entering (2.2).

The next theorem shows point identification under either of the three assumptions.

Theorem 4.2. Under the assumptions of Theorem 3.1 replacing (A2) by (A2'), $F_{Y_0,Y_1|\overline{\Gamma}_{2-L},X}$ is identified under Assumption 4.1, while $F_{Y_0,Y_1|\overline{\Theta},\overline{\Gamma}_{2-L},X}$ is identified under Assumption 4.2 or 4.3.

The following result is consequence Theorem 3.2. Carneiro, Hansen and Heckman (2003) and Cunha and Heckman (2007) rely on a factor model to obtain related results.

Theorem 4.3. Suppose (2.2) corresponds to model RCR. Under the assumptions of Theorem 4.2 with Assumption 4.2 or 4.3, and Assumption 3.1, $F_{\Delta,\mathbb{E}[\Delta|\mathcal{I}]|X}$ is identified. Moreover, if $\mathbb{E}[\Delta|\mathcal{I}]$ is not $\sigma(\overline{\Gamma}_{2-L},X)$-measurable, then $F_{\mathrm{EAR},\mathrm{EPR}|Z,X}$ is identified.

UCATE is a building block to obtain effects that depend on averages. The structural parameter that we now introduce plays a similar role, but allows to obtain all effects that depend on $F_{Y_0,Y_1|X}$. It is clearly identified from Theorem 4.2 and policy invariant.

Definition 4.1. The Unobservables Conditioned Distribution of Treatment Effects is defined, for δ in $\mathbb{R}$ and (θ,γ_{2-L},x) in $\mathrm{supp}(\mathbb{P}_{\overline{\Theta},\overline{\Gamma}_{2-L},X})$, as

$$\mathrm{UCDITE}(\delta,\theta,\gamma_{2-L},x) := f_{\Delta|\overline{\Theta},\overline{\Gamma}_{2-L},X}(\delta|\theta,\gamma_{2-L},x) .$$

UCDITE is a key quantity to obtain all effects that depend on $F_{Y_0,Y_1|X}$, for example

$$\mathbb{P}(\Delta > 0|X = x) = \int_{\mathbb{R}} \mathbb{1}\{\delta > 0\} \int_{\mathbb{R}^L} \mathrm{UCDITE}(\delta,\theta,\gamma_{2-L},x) d\mathbb{P}_{\overline{\Theta},\overline{\Gamma}_{2-L}|X}(\theta,\gamma_{2-L}|x)$$

$$f_{Y_0,Y_1|X}(y_0,y_1|x) = \int_{\mathbb{R}^L} \mathrm{UCDITE}(y_1 - y_0,\theta,\gamma_{2-L},x) d\mathbb{P}_{Y_0,\overline{\Theta},\overline{\Gamma}_{2-L}|X}(y_0,\theta,\gamma_{2-L}|x)$$

$$f_{\Delta|D=1,Y_0,X}(\delta|y_0,x) = \int_{\mathbb{R}^L} h_{\mathrm{TT}}(\theta,\gamma_{2-L},x) \mathrm{UCDITE}(\delta,\theta,\gamma_{2-L},x) d\mathbb{P}_{Y_0,\overline{\Theta},\overline{\Gamma}_{2-L}|X}(y_0,\theta,\gamma_{2-L}|x) .$$

Appendix 1: Modifications When Some of the Instruments are Discrete

Because of its importance in applications, we consider the case where some instruments are discrete. For simplicity, we only detail the case with one additional instrument B which is binary and use (A4). We replace (2.2) by

$$(4.1) \qquad D = \mathbb{1}\left\{\Theta - \alpha B - \Gamma_1(Z_1 + g(Z_{2-L},B)) - \sum_{l=2}^{L} \Gamma_l f_l(Z_l,B) > 0\right\}$$

where $\overline{\alpha} := \alpha/\Gamma_1$, $g(Z_{2-L}, B)$ and $f_l(Z_l, B)$ are integrable. Replace (A2), (A2'), (A4) by

(AB2) $(Z, B) \perp (Y_d, \overline{\Theta}, \overline{\Gamma}_{2-L}) \,|X$ for $d = 0, 1$;

(AB4) $\forall (x, z_{2-L}, b) \in \text{supp}\left(\mathbb{P}_{X, Z_{2-L}, B}\right),$

$\qquad \text{supp}\left(\mathbb{P}_{Z_1|Z_{2-L}, B, X}(\cdot|z_{2-L}, b, x)\right) \supseteq \text{supp}\left(\mathbb{P}_{\overline{\Theta} - \overline{\alpha}b - g(z_{2-L}, b) - \sum_{l=2}^{L} \overline{\Gamma}_l f_l(z_l, b)|X}(\cdot|x)\right);$

(N1)-(N4) by (NB1)-(NB4) where we work with $f_l(Z_l, B)$ for $l = 2, \ldots, L$, $f_l(Z_l)$ and $g(Z_{2-L}, B)$; and the conditioning on X by conditioning on X and B.

Based on (AB2), (AB4), (NB1)-(NB4), the functions g and f_l for $l = 2, \ldots, L$ are identified and we can assume that the first-stage equation is

$$(4.2) \qquad D = \mathbb{1}\left\{-\overline{\Theta} + \overline{\alpha}B + \sum_{l=2}^{L} \overline{\Gamma}_l \widetilde{Z}_L < \widetilde{Z}_1\right\}$$

where $\widetilde{Z}_1 := -Z_1 + g(Z_{2-L}, B)$ and $\widetilde{Z}_l := f_l(Z_l, B)$ for $l = 2, \ldots, L$. We make Assumptions 2.4 and 2.5 on $\text{supp}(\mathbb{P}_{Z_{2-L}|B, X})$ where we also condition on B.

Under such assumptions, Theorem 3.1 holds with $\mathbb{P}_{Y_d, \overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X}$ for $d = 0, 1$ and $b = 0, 1$. Thus, UCATE and UCQTE are identified when we replace conditioning on $\overline{\Theta}, \overline{\Gamma}_{2-L}, X$ by $\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X$ for $(\theta, \gamma_{2-L}, b, x)$ in $\text{supp}\left(\mathbb{P}_{\overline{\Theta} + \overline{\alpha}B, B, \overline{\Gamma}_{2-L}|B, X}\right)$. This gives, for $b = 0, 1$, two formulas for the effects which are weighted integrals.

UCDITE is identified when we replace (AB2) by $(Z, B) \perp (Y_0, Y_1, \overline{\Theta}, \overline{\Gamma}_{2-L}) \,|X$, Assumption 4.2 by $Y_0 \perp Y_1 \,|\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X$, and Assumption 4.3 (i) by $Y_0 \perp \Delta \,|\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X$. Without appealing to such assumptions, we obtain intersection bounds, for example, for x in $\text{supp}(\mathbb{P}_X)$ and $(y_0, y_1) \in \mathbb{R}^2$,

$$F^L_{Y_0, Y_1|X}(y_0, y_1|x) \leq F_{Y_0, Y_1|X}(y_0, y_1|x) \leq F^U_{Y_0, Y_1|X}(y_0, y_1|x)$$

where, defining $F^L_{Y_0, Y_1|\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X}(y_0, y_1|\theta, \gamma_{2-L}, x)$ and $F^U_{Y_0, Y_1|\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X}(y_0, y_1|\theta, \gamma_{2-L}, x)$ like $F^L_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0, y_1|\theta, \gamma_{2-L}, x)$ and $F^U_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0, y_1|\theta, \gamma_{2-L}, x)$,

$$F^L_{Y_0, Y_1}(y_0, y_1|x) := \max_{b=0,1} \int_{\mathbb{R}^L} F^L_{Y_0, Y_1|\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}, X}(y_0, y_1|\theta, \gamma_{2-L}, x) d\mathbb{P}_{\overline{\Theta} + \overline{\alpha}b, \overline{\Gamma}_{2-L}|X}(\theta, \gamma_{2-L}|x)$$

$$F_{Y_0,Y_1}^U(y_0, y_1|x) := \min_{b=0,1} \int_{\mathbb{R}^U} F_{Y_0,Y_1|\overline{\Theta}+\overline{\alpha}b,\overline{\Gamma}_{2-L},X}^U(y_0, y_1|\theta, \gamma_{2-L}, x) d\mathbb{P}_{\overline{\Theta}+\overline{\alpha}b,\overline{\Gamma}_{2-L}|X}(\theta, \gamma_{2-L}|x) .$$

Alternatively, effects conditional on $\{\overline{\Theta} = \theta, \overline{\alpha} = a, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\}$ are identified under Assumption 4.3, replacing Y_0, Δ by $\overline{\Theta}, \overline{\alpha}$ and without conditioning on $\overline{\Theta}$.

APPENDIX 2: VARIANCE OF TREATMENT EFFECTS

We now assume that $\mathbb{E}\left[Y_0^2 + Y_1^2\right] < \infty$, take $(\theta, \gamma_{2-L}, x) \in \text{supp}(\mathbb{P}_{\overline{\Theta},\overline{\Gamma}_{2-L},X})$, and use

$$V_1^U(\theta, \gamma_{2-L}, x) := 2\left(\mathbb{E}\left[Y_0^2 \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right] + \mathbb{E}\left[Y_1^2 \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right]\right)$$

$$V_2^U(\theta, \gamma_{2-L}, x) := \mathbb{E}\left[Y_0^2 \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right] + \mathbb{E}\left[Y_1^2 \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right]$$
$$- 2\mathbb{E}\left[Y_0 \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right] \mathbb{E}\left[Y_1 \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right] .$$

Under the assumptions of Theorem 3.1 and assuming

$$\mathbb{E}[(Y_0 - \mathbb{E}[Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X])(Y_1 - \mathbb{E}[Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X])|\overline{\Theta}, \overline{\Gamma}_{2-L}, X] \geq 0 \ a.s.$$

for (4.3) below, we easily obtain that

$$Var\left(\Delta \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right) \leq V_1^U(\theta, \gamma_{2-L}, x)$$

(4.3) $$Var\left(\Delta \middle| \overline{\Theta} = \theta, \overline{\Gamma}_{2-L} = \gamma_{2-L}, X = x\right) \leq V_2^U(\theta, \gamma_{2-L}, x)$$

These upper bounds yield unconditional bounds when taking expectation of the above inequalities evaluated at the random vectors.

Based on the fact conditional variance of the sum of two uncorrelated variables is the sum of the conditional variances, replacing Assumption 4.3 by the weaker assumption

$$\mathbb{E}[(Y_0 - \mathbb{E}[Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X])(\Delta - \mathbb{E}[\Delta|\overline{\Theta}, \overline{\Gamma}_{2-L}, X])|\overline{\Theta}, \overline{\Gamma}_{2-L}, X] = 0 \ a.s.$$

yields

$$Var(\Delta|\overline{\Theta}, \overline{\Gamma}_{2-L}, X) = Var(Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X) - Var(Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X) .$$

Appendix 3: Proofs

Proof of Theorem 2.1. We have $\mathbb{E}[W|X] = -\mathbb{E}[\overline{\Theta}|X]$, while for $l = 2, \ldots, L$, $\mathbb{E}[W|Z_l, X] = -\mathbb{E}[\overline{\Theta}|X] + \mathbb{E}[\overline{\Gamma}_l|X]f_l(Z_l)$, thus $f_l(Z_l) = \mathbb{E}[W|Z_l, X] - \mathbb{E}[W|X]$. This identifies the functions f_l on $\mathrm{supp}(\mathbb{P}_{Z_l})$ for $l = 2, \ldots, L$. Finally, g is identified using

$$g(Z_{2-L}) = \mathbb{E}[W|Z_{2-L}, X] - \mathbb{E}[W|X] - \sum_{l=2}^{L} f_l(Z_l) .$$

Proof of Theorem 3.1. (A2) implies that for $(z, x) \in \mathrm{supp}(\mathbb{P}_{Z,X})$, $y_1 \in \mathbb{R}$

$$\mathbb{E}\left[\mathbb{1}\{Y \leq y_1\}D \big| \overline{Z} = z, X = x\right] = \mathbb{E}\left[\mathbb{1}\{Y_1 \leq y_1\}D \big| \overline{Z} = z, X = x\right]$$

$$= \mathbb{E}\left[\mathbb{1}\{Y_1 \leq y_1\}\mathbb{1}\left\{-\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l z_l < z_1\right\} \bigg| X = x\right] .$$

Using (A3) and (A4), this yields that $F_{Y_1, -\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l z_l|X}(\cdot|x)$ is identified. Thus, for every t, s in $\mathbb{R}$, $x \in \mathrm{supp}(\mathbb{P}_X)$, and z_{2-L} in $\mathrm{supp}(\mathbb{P}_{Z_{2-L}|X}(\cdot|x))$, $\mathbb{E}\left[\exp\left(itY_1 - is\overline{\Theta} + is\sum_{l=2}^{L} \overline{\Gamma}_l z_l\right) \big| X = x\right]$ is identified. Under Assumption 2.5, (A1) implies that we can extend as an entire function $z_{2-L} \to \mathbb{E}\left[\exp\left(itY_1 - is\overline{\Theta} + is\sum_{l=2}^{L} \overline{\Gamma}_l z_l\right) \big| X = x\right]$ and (S2) that for every $t, s \in \mathbb{R}$ and $z_{2-L} \in \mathbb{R}$, $\mathbb{E}\left[\exp\left(itY_1 - is\overline{\Theta} + is\sum_{l=2}^{L} \overline{\Gamma}_l z_l\right) \big| X = x\right]$ is identified. Thus, $(t, s, z_{2-L}) \in \mathbb{R}^2 \times (\mathbb{R} \setminus \{0\})^{L-1} \to \mathbb{E}\left[\exp\left(itY_1 + is\overline{\Theta} + i\sum_{l=2}^{L} \overline{\Gamma}_l z_l\right) \big| X = x\right]$ is identified. We obtain directly this result if we make Assumption 2.4 instead of Assumption 2.5. Now, by continuity, the Fourier transform of $\mathbb{P}_{Y_1, \overline{\Theta}, \overline{\Gamma}_{2-L}|X}(\cdot|x)$ is identified everywhere. The injectivity of the Fourier transform of measures allows to conclude.

If we maintain (A4') instead of (A4), the above argument yields that $\mathbb{P}_{Y_d, -\overline{\Theta} + \sum_{l=2}^{L} \overline{\Gamma}_l z_l}$ for $j = 0, 1$ is identified on $\mathbb{R} \times \mathrm{supp}(\mathbb{P}_{Z_1|Z_{2-L}, X}(\cdot|z_{2-L}, x))$ for every (z_{2-L}, x) in $\mathrm{supp}(\mathbb{P}_{Z_{2-L}, X})$. (A4'b) and (A4'c) now implies that, for every t, s in $\mathbb{R}$, (z_{2-L}, x) in $\mathrm{supp}\left(\mathbb{P}_{Z_{2-L}, X}\right)$ and z_1 in $\mathrm{supp}\left(\mathbb{P}_{Z_1|\overline{Z}_{2-L}, X}(\cdot|z_{2-L}, x)\right)$, $\mathbb{E}\left[\exp\left(itY_1 - is\overline{\Theta} + iz_1\sum_{l=2}^{L} \overline{\Gamma}_l z_l\right) \big| X = x\right]$ is identified. Now, we can either conclude under Assumption 2.4 or Assumption 2.5.

Identification of $F_{Y_0, \overline{\Theta}, \overline{\Gamma}_{2-L}}$ is obtained in the same way replacing D by $1 - D$.

Proof of Theorem 3.2. This follows from the fact that, based on (3.1)-(3.2), we have

$$\mathbb{E}[Y_1 - Y_0|\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X] = \mathbb{E}\left[\mathbb{E}[Y_1 - Y_0|\mathcal{I}]\middle|\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X\right]$$

$$= \mathbb{E}\left[\mathbb{E}[Y_1 - Y_0|\overline{\Gamma}_{2-L}, X] + \eta_1 - \eta_0\middle|\overline{\Gamma}_{2-L}, \eta_1 - \eta_0, X\right]$$

$$= \mathbb{E}[Y_1 - Y_0|\overline{\Gamma}_{2-L}, X] + \eta_1 - \eta_0 = \mathbb{E}[\Delta|\mathcal{I}] .$$

Now, because $\mathbb{E}[\Delta|\mathcal{I}]$ is not $\sigma(\overline{\Gamma}_{2-L}, X)$-measurable and we assume (U2), for every $(\gamma_{2-L}, x) \in \text{supp}(\mathbb{P}_X)$, there are two draws $\theta_1 \neq \theta_2$ from $F_{\overline{\Theta}|\overline{\Gamma}_{2-L}, X}(\cdot|\gamma_{2-L}, x)$ such that

$$\Gamma_1(\gamma_{2-L}, x)(\theta_1 - \theta_2) = \text{UCATE}(\theta_1, \gamma_{2-L}, x) - \text{UCATE}(\theta_2, \gamma_{2-L}, x) \neq 0 .$$

Thus $\Gamma_1(\gamma_{2-L}, x)$ and $\Gamma_0(\gamma_{2-L}, x) = \text{UCATE}(\theta_1, \gamma_{2-L}, x) - \Gamma_1(\gamma_{2-L}, x)\theta_1$ are identified. Stated otherwise, conditioning on a fixed value of $(\overline{\Gamma}_{2-L}, X)$, Γ_0 and Γ_1 are constant and parameters in a regression without error. Thus EAR is uniquely determined once we condition on a fixed value of $(\overline{\Theta}, \overline{\Gamma}_{2-L}, X)$ and Z. This yields the result by integration.

Proof of Theorem 4.1. The first inequality is based on the Frechet-Hoeffding bounds applied to the conditional CDFs which are CDFs as well. The second one is a consequence of the Makarov bounds (Makarov (1981)) for the same reason. Sharpness is discussed in Firpo and Ridder (2008). The last two inequalities are obtained by simple manipulations.

Proof of Corollary 4.1. The bounds are obtained by taking expectations of the inequalities in Theorem 4.1 evaluated at the random vectors.

Proof of Theorem 4.2. Assumption 4.1 implies that

$$F_{Y_0, Y_1|\overline{\Gamma}_{2-L}, X}(y_1, y_0|\gamma_{2-L}, x) = F_{Y_0|\overline{\Gamma}_{2-L}, X}(y_0|\gamma_{2-L}, x)F_{Y_1|\overline{\Gamma}_{2-L}, X}(y_1|\gamma_{2-L}, x) ,$$

similarly, Assumption 4.2 implies that

$$F_{Y_0, Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0, y_1|\theta, \gamma_{2-L}, x) = F_{Y_0|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_0|\theta, \gamma_{2-L}, x)F_{Y_1|\overline{\Theta}, \overline{\Gamma}_{2-L}, X}(y_1|\theta, \gamma_{2-L}, x)$$

and the right-hand sides are identified.

Let us now consider Assumption 4.3. Based on Theorem 3.1, $F_{Y_0|\overline{\Theta},\overline{\Gamma}_{2-L},X}(\cdot|\theta,\gamma_{2-L},x)$ and $F_{Y_0+\Delta|\overline{\Theta},\overline{\Gamma}_{2-L},X}(\cdot|\theta,\gamma_{2-L},x)$ are identified. Estimating $F_{\Delta|\overline{\Theta},\overline{\Gamma}_{2-L},X}(\cdot|\theta,\gamma_{2-L},x)$ and thus $F_{Y_0,Y_1|\overline{\Theta},\overline{\Gamma}_{2-L},X}(\cdot|\theta,\gamma_{2-L},x)$ is a deconvolution problem under (i).

Proof of Theorem 4.3. This is a consequence of theorems 4.2 (and its proof) and 3.2.

References

[1] ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings". *Econometrica*, **70**, 91–117.

[2] ABBRING, J. H., AND J. J. HECKMAN (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation". *Handbook of Econometrics*, J.J. Heckman and E.E. Leamer (eds.), Vol. 6, North Holland, Chapter 72.

[3] BERAN, R., AND P. W. MILLAR (1994): "Minimum Distance Estimation in Random Coefficients Regression Models". *Annals of Statistics*, **22**, 1976–1992.

[4] BERAN, R., A. FEUERVERGER, AND P. HALL (1996): "On Nonparametric Estimation of Intercept and Slope in Random Coefficients Regression". *Annals of Statistics*, **24**, 2569–2592.

[5] BJÖRKLUND, A., AND R. MOFFITT (1987): "The Estimation of Wage and Welfare Gains in Self-Selection Models", *Review of Economics and Statistics*, **69**, 42–49.

[6] CARNEIRO, P., K. T. HANSEN, AND J. HECKMAN (2003): "Estimating Distributions of Treatment Effects With an Application to the Return to Schooling and Measurement of the Effect of Uncertainty on College Choice". *International Economic Review*, **44**, 361–422.

[7] CARNEIRO, P., J. HECKMAN, AND E. VYTLACIL (2010): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin". *Econometrica*, **78**, 377–394.

[8] CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects". *Econometrica*, **73**, 245–261.

[9] CUNHA, F., AND J. HECKMAN (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling". *Labour Economics*, **14**, 870–893.

[10] DEVROYE, L. (1989): "Consistent Deconvolution in Density Estimation". *The Canadian Journal of Statistics*, **17**, 235–239.

[11] FAN, Y., E. GUERRE AND D. ZHU (2014): "Partial Identification and Confidence Sets for Functional of the Joint Distribution of Potential Outcomes". Working paper.

[12] FAN, Y., AND S. S. PARK (2010): "Sharp Bounds on the Distribution of Treatment Effects and Their Statistical Inference". *Econometric Theory*, **26**, 931–951.

[13] FIRPO, S., AND G. RIDDER (2008): "Bounds on Functionals of the Distribution of Treatment Effects". Working paper.

[14] FOX, J., AND A. GANDHI (2011): "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models". Working Paper.

[15] GAILLAC, C., AND E. GAUTIER (2015): "Estimation of the Distribution of Random Coefficients with Bounded Regressors". Working paper.

[16] GAUTIER, E., AND Y. KITAMURA (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models". *Econometrica*, **81**, 581–607.

[17] HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts". *Review of Economic Studies*, **64**, 487–635.

[18] HECKMAN, J. J., AND E. VYTLACIL (2001) "Policy Relevant Treatment Effects". *American Economic Review Papers and Proceedings*, **91**, 107–111.

[19] HECKMAN, J. J., AND E. VYTLACIL (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation". *Econometrica*, **73**, 669–738.

[20] HODERLEIN, S., J. KLEMELÄ, AND E. MAMMEN (2010): "Analyzing the Random Coefficient Model Nonparametrically". *Econometric Theory*, **26**, 804–837.

[21] ICHIMURA, H., AND T. S. THOMPSON (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution". *Journal of Econometrics*, **86**, 269–295.

[22] IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects". *Econometrica*, **62**, 467–475.

[23] IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity Corresponding". *Econometrica*, **77**, 1481–1512.

[24] KASY, M. (2011): "Identification in Triangular Systems Using Control Functions". *Econometric Theory*, **27**, 663–671.

[25] KITAGAWA, T. (2009): "Identification Region of the Potential Outcome Distributions under Instrument Independence". Working paper.

[26] KLEIN, T. (2010): "Heterogeneous Treatment Effects: Instrumental Variables Without Monotonicity?". *Journal of Econometrics*, **155**, 99–116.

[27] LEWBEL, A. (2007): "Endogenous Selection or Treatment Model Estimation". *Journal of Econometrics*, **141**, 777–806.

[28] LEWBEL, A. AND K. PENDAKUR (2012): "Generalized Random Coefficients With Equivalence Scale Applications". Working paper.

[29] Makarov, G. D. (1981): "Estimates of the Distribution Function of a Sum of Two Random Variables when the Marginal Distributions are Fixed". *Theory of Probability and its Applicatons*, **26**, 803–806.

[30] Newey, W. K., Powell, J. L. and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models". *Econometrica*, **67**, 565–603.

[31] Small and Tan (2007): "A Stochastic Monotonicity Assumption for the Instrumental Variables Method". Working paper.

[32] Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result". *Econometrica*, **70**, 331–341.

Toulouse School of Economics, Université Toulouse 1 Capitole, 21 allée de Brienne, 31000 Toulouse, France.

E-mail address: Eric.Gautier@tse-fr.eu

Boston College, Chestnut Hill, MA 02467, USA.

E-mail address: Stefan.Hoderlein@bc.edu