Testing and Relaxing the Exclusion Restriction in the Control Function Approach

Xavier D'Haultfœuille* Stefan Hoderlein† Yuya Sasaki‡

CREST-ENSAE Emory Vanderbilt

July 1, 2020

Abstract

The control function approach which employs an instrumental variable excluded from the outcome equation is a very common solution to deal with the problem of endogeneity in nonseparable models. Exclusion restrictions, however, are frequently controversial. We first argue that, in a nonparametric triangular structure typical of the control function literature, one can actually test this exclusion restriction provided the instrument satisfies a local irrelevance condition. Second, we investigate identification without such exclusion restrictions, i.e., if the "instrument" that is independent of the unobservables in the outcome equation also directly affects the outcome variable. In particular, we show that identification of average causal effects can be achieved in the two most common special cases of the general nonseparable model: linear random coefficients models and single index models.

Keywords: identification, control function, endogenous regressors, exclusion restriction, nonseparable models.

JEL Codes: C14, C31, C36.

* Xavier D'Haultfœuille: CREST-ENSAE, 5 avenue Henry Le Chatelier 91120 Palaiseau, France, email: xavier.dhaultfoeuille@ensae.fr.

† Stefan Hoderlein: Emory University, Department of Economics, Atlanta, GA 30322-2240 , USA, email: stefan.hoderlein@emory.edu.

‡ Yuya Sasaki: Department of Economics, Vanderbilt University, VU Station B, Box #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA, email: yuya.sasaki@vanderbilt.edu.

1 Introduction

The control function approach is a popular way to use instrumental variables (IV) in nonlinear models with endogeneity. An important reason is that, as Imbens and Newey (2009) demonstrated in a seminal paper, it applies to a large class of models, including nonseparable, nonparametric ones with possibly multi-dimensional unobservables. Like most common IV methods, the control function approach crucially relies on the exclusion restriction that the IV is not a part of the structural outcome equation. Exclusion restrictions, however, are frequently controversial. The aim of this paper is to show that, in certain circumstances, the conditions usually imposed in the control function approach make it possible to test this exclusion restriction as well as to relax it.

Specifically, suppose that the endogenous variable X is continuous and an additional condition on the instrument Z, which we term "local irrelevance condition", holds. In this context, we derive a testable implication for the validity of the exclusion restriction. Roughly speaking, the local irrelevance condition, which is itself testable, requires that for a subset of individuals, possibly of measure zero, a change in Z does not affect their X. This condition does not contradict the usual relevance conditions on the IV if, basically, Z has an heterogeneous effect on X.[1] We develop a bootstrap test for our testable implication, and show its size control and consistency.

Second, we show that average causal effects of X and Z on Y may be identified without exclusion restrictions. We do so in what are arguably the two most common specifications of the general nonseparable model analyzed in Imbens and Newey (2009): random coefficients models and single index models. Remarkably, in both cases, point identification of some causal effects of interest already holds with a discrete instrument. A richer support of Z then leads to identification of additional causal effects.

Though the arguments differ in important aspects from one model to another, the general identifying strategy is the same in both cases. In a first step, we exploit the local irrelevance condition

[1] We show in particular that in a (generalized) location-scale model on X, this holds if and only if the model is heteroskedastic in Z, or the location function is not one-to-one.

to identify the direct effect of Z on Y. Intuitively, if the instrument is locally irrelevant, there is a subpopulation for which a change in Z will affect Y only directly, and not indirectly through the effect of Z on X. In a second step, we consider other subpopulations for which Z has an effect on X. For such subpopulations, the change in Y is due to the direct effect of Z but also to its indirect effect, i.e., through X. Because we have already identified the direct effect of Z in step one, we can recover the indirect effect and thus, at the end, the causal effect of X on Y. While the arguments used to establish identification in these models share this general strategy, there are also profound differences between models, entailing separate formal identification results across the models.

Related Literature The control function approach has a long tradition in econometrics since, at least, the work of Heckman (1979) (see Wooldridge, 2015, for a recent survey). Historically, this approach has mainly relied on two sets of restrictions: 1) functional-form restrictions and 2) exclusion restrictions. As Imbens and Newey (2009) made clear, functional-form restrictions are actually superfluous for the control function approach to work. Exclusion restrictions, on the other hand, still remain essential in their framework. The same is true for the rest of the control function literature, with the exception of Klein and Vella (2010), which we discuss below. The models of Chesher (2003), Vytlacil and Yildiz (2007), Hoderlein and Sasaki (2013), D'Haultfoeuille and Février (2015) and Torgovitsky (2015), to cite but a few, all feature exclusion restrictions, and are concerned with other parts of the model specification.

However, the exclusion restriction is controversial. In their critique of natural experiments, Rosenzweig and Wolpin (2000) discuss several important examples where the instrument, while exogenous in our sense, may have a direct effect on the outcome variable. Also, van den Berg (2007) considers randomized controlled trials for which some time elapses between the moment the agents realizes she may be treated and the moment when the treatment takes place. In such common situations, the agent has an incentive to learn the value of the instrument. Then, van den Berg shows that the exclusion restriction can be violated if the interaction between the effort of the agent and the treatment affects the outcome variable. In the binary treatment case, Jones (2015) presents several examples of economic models where the exclusion restriction is likely to be violated, especially among inframarginal agents such as always- and never-takers. In the very typical analyses of returns to educations where college subsidy is used as an instrument, for example, the college subsidy can generate an income effect for always takers and thus may affect labor market outcomes (Jones, 2015, Section 3.1).[2]

Thus, our paper contributes to the literature by first showing that under some natural and testable restrictions on the distribution of (X, Z), the exclusion restriction can be tested within the control function approach. We are not the first, however, to investigate testability without excluding instruments in models with endogenous variables. In the context of linear IV models, Bound and Jaeger (2000) and Altonji et al. (2005) suggest a test of the exclusion restriction based on a similar idea as ours. Specifically, if the first-stage effect is zero for subgroups defined by covariates, the effect of Z in the reduced form equation should be zero as well. The zero first-stage effect is close in spirit to our local irrelevance condition, but in our context subgroups are constructed with X and Z only. Also, our test works beyond linear models – actually, it does not rely on any functional form restrictions on the structural equation, though it does impose conditions (monotonicity) on the first stage.

The related test of Kitagawa (2015) does not impose any functional forms either, and may also be seen as a test of the exclusion restriction if we maintain exogeneity and a monotonicity condition. The two procedures have also important differences. First and foremost, the test of Kitagawa (2015) applies to the framework of Imbens and Angrist (1994) with binary treatment and binary instrument, whereas we consider the case of continuous treatment, with either discrete or continuous instrument. Second, the monotoncity restrictions are different in the two settings. Caetano et al. (2016) also develop a nonparametric test of the exclusion restriction,[3] but under a different condition on (X, Z). Specifically, they assume that X admits a mass point at the boundary of its support, whereas we leverage on the aforementioned local irrelevance condition.

[2]It is easy to see that these examples would generalize to situations with continuous X, as analyzed in this paper.

[3]Caetano et al. (2016) frame their test as a test of the independence and monotonicity conditions behind the control function approach, while maintaining the exclusion restriction. But if, as we do here, these independence and monotonicity conditions are maintained, their procedure becomes a test of the exclusion restriction.

Turning to identification, several papers have shown how to recover causal effects without exclusion restrictions in linear models. In particular, van Kippersluis and Rietveld (2018) show that if we assume homogeneous effects across subgroups and we have a zero first stage effect, we can identify both the effects of X and Z. Their idea is related to our main identification idea for the random coefficients models. The main difference is that we allow for heterogeneous treatment effects in this model.[4] Still in linear models, Rigobon (2003) and Lewbel (2012) show that second-order moment conditions have enough identifying power in systems of simultaneous equations, provided the model displays some heteroskedasticity. Klein and Vella (2010), who rely on a control function specification, exploit heteroskedasticity as well. However, their approach crucially hinges on the linearity of the structural and first-stage equations, whereas we establish identification in possibly nonlinear or nonparametric models. Our paper is also related to recent papers showing identification in nonseparable models when instruments has limited support. Newey and Stouli (2018, 2019) show that under restrictions on the structural functions, including random coefficients models similar to that considered below, identification can be achieved with a discrete instrument (see also Masten and Torgovitsky, 2016, for a similar result). While we consider less general effects than them, our approach does not require any exclusion restriction. Finally, our paper is related to Feng (2020) in that both exploit variations in covariates to identify causal effects. But they are quite different otherwise, as Feng (2020) considers a discrete X and still relies on exclusion restrictions.

2 The General Model

We introduce in this section the general class of triangular models that we discuss throughout this paper. The class of models is formally defined through the following system of equations:

$$\begin{cases} Y &= g(X, Z, \varepsilon) \\ X &= h(Z, \eta) \end{cases} \tag{2.1}$$

[4] Our identification strategies based on index restrictions are different and exploit heterogeneous responses in the first stage, which linear models cannot leverage.

where, for simplicity, we assume that both X and Z are scalar variables. We could allow both equations to depend in addition on a random vector of exogenous regressors denoted by S. In line with the treatment effect literature, we omit this dependence, as the analysis can be done conditionally on $S = s$, for any values of s in the support of S. In addition to (2.1), we impose a few regularity conditions on (X, Y, Z) that are summarized in the following assumption. Hereafter, for any random variable A, we denote by $\text{Supp}(A)$ and F_A the support and the cumulative distribution function (cdf for short) of A. Similarly, for any other random variable B, $\text{Supp}(A|B = b)$ and $F_{A|B}(\cdot|b)$ denote the conditional support and conditional cdf of A given $B = b$.

Assumption 1. *The model is defined by (2.1), where the random vector* $(X, Y, Z) : \Omega \to Supp(X, Y, Z) \subseteq \mathbb{R}^3$ *is observed, the random variables* $\varepsilon : \Omega \to \mathcal{E}$ *and* $\eta : \Omega \to \mathbb{R}$ *are unobserved, and all random variables are defined on a complete probability space* $(\Omega, \mathcal{F}, P)$. *There exist regular conditional probabilities* $\Pr(\varepsilon \in \cdot\ |\eta = u)$ *and* $\Pr(Y \in \cdot\ |X = x, Z = z)$ *that are continuous in* u *and* (x, z)*, respectively. The distribution of* X *admits a density with respect to the Lebesgue measure.*

Importantly, the support $\mathcal{Z}$ of Z can be a strict subset of $\mathbb{R}$. In particular, Z may be binary, though in some cases considered later, point identification will require Z to be continuous, though not (necessarily) with a large support. Note also that the outcome Y may be continuous or discrete, and that ε is allowed to be (countably) infinite dimensional – in fact, it may be an element of an even more general space, but we desist here from this unnecessary generality.

Additional assumptions are needed to identify meaningful objects. The first is the following exogeneity of Z.

Assumption 2 (Exogeneity)**.** $Z \perp\!\!\!\perp (\varepsilon, \eta)$.

This assumption states that the instrument is jointly independent from all unobservables in the system. It is very commonly assumed on instrumental variables in the literature in nonseparable models – see Chesher (2003), Imbens and Newey (2009), Hoderlein and Sasaki (2013), D'Haultfoeuille and Février (2015) and Torgovitsky (2015), to cite but a few. Importantly, this assumption is less restrictive in our scenario as Z may directly affect the outcome.

Our second assumption – equally typical of the control function literature – specifies the way the instrument enters the first stage equation.

Assumption 3 (First-Stage Monotonicity). *$h(z, \cdot)$ is strictly increasing for every $z \in \mathcal{Z}$, and $\eta \sim \mathcal{U}[0, 1]$.*

Similarly to the exogeneity condition, this assumption is very common in papers relying on control functions – see, among others, Imbens and Newey (2009), D'Haultfoeuille and Février (2015), Torgovitsky (2015) and Caetano et al. (2016). While it allows for general, nonparametric responses to changes in Z, it rules out more general forms of heterogeneity in the first stage, e.g., vectors of random coefficients – see Gautier and Hoderlein (2015) and Hoderlein et al. (2017) for example. Given the monotonicity of $h(z, \cdot)$, the uniform distribution condition on η is a mere normalization as soon as F_η is continuous. Together with Assumption 2, it implies that η is identified by $\eta = F_{X|Z}(X|Z)$.

3 Testing the exclusion restriction

3.1 Testable implications

In many applications, a candidate variable Z that *may* satisfy the exclusion restriction $g(X, Z, \varepsilon) = g(X, \varepsilon)$ is available, but it is uncertain whether it satisfies it or not. In this section, we establish that the exclusion restriction is actually testable under the local irrelevance condition below.

Assumption 4 (Locally Irrelevant Instrument). *There exists $(x^*, z, z') \in \mathbb{R} \times \mathcal{Z}^2$ such that $z \neq z'$ and $F_{X|Z}(x^*|z) = F_{X|Z}(x^*|z') \in (0, 1)$.*

This assumption may hold even if Z is binary. Also, since it only involves observed variables, the condition is testable. Similar assumptions are imposed by Torgovitsky (2015) and D'Haultfoeuille and Février (2015) to point identify g under the additional restrictions that g does not depend on Z and is strictly increasing in ε. This condition holds in the "degenerate" case where Z is independent of X. Otherwise, Assumption 4 holds if, basically, Z has heterogeneous effects on X. To see this, note that under Assumption 3, Assumption 4 is equivalent to assuming the existence of $(z, z', u) \in \mathcal{Z}^2 \times (0,1)$ such that $h(z,u) = h(z',u)$. Hence, while some individuals may be affected by a change in Z with η kept constant, units with $\eta = F_{X|Z}(x^*|z)$ are not affected by a shift of Z from z to z'.

The following examples illustrate this idea that Assumption 4 requires heterogeneous effects of Z on X. In particular, in the first example, heteroskedasticity or some non-monotonicity is necessary and sufficient for Assumption 4 to hold.

Example 1. *Suppose that Assumptions 1-2 hold and assume that*

$$X = \mu\left(\psi(Z) + \sigma(Z)\eta\right), \tag{3.1}$$

where $\mu(\cdot)$ is strictly increasing, $\sigma(Z) > 0$ and Supp$(\eta) = \mathbb{R}$.[5] *In this generalized location-scale model, Assumption 4 holds if and only if*

- *either $\sigma(\cdot)$ is constant and there exists (z, z'), $z \neq z'$, such that $\psi(z) = \psi(z')$;*
- *or $\sigma(\cdot)$ is not constant.*

See Appendix C.1 for a proof.

Example 2. *Suppose that Assumptions 1-2 hold and assume that there exist $P(\cdot)$ and $\pi(\cdot)$ taking values in $\mathbb{R}^k$ $(k \geq 1)$ and a cdf F such that for all $(x,z) \in$ Supp$(X|Z=z)$,*

$$F_{X|Z}(x|z) = F\left(P(z)'\pi(x)\right).$$

This corresponds to the distribution regression introduced by Foresi and Peracchi (1995). In this model, Assumption 4 holds if and only if there exists (x^*, z, z'), $z \neq z'$, such that

$$(P(z) - P(z'))'\pi(x^*) = 0 \tag{3.2}$$

and $F\left(P(z)'\pi(x)\right) \in (0,1)$. In particular, if $Z \in \{0,1\}$ (or the effect of Z is linear), so that $P(z) = (1,z)'$ and $\pi(x) = (\pi_0(x), \pi_1(x))'$, Assumption 4 holds if and only if there exists x^* such that $\pi_1(x^*) = 0$

[5]In Model (3.1), η is not uniform, but $\eta' = F_\eta(\eta)$ would in $X = \mu\left(\psi(Z) + \sigma(Z)F_\eta^{-1}(\eta')\right)$. We consider (3.1) rather than this latter form to ease the connection with location-scale models.

and $F_{X|Z}(x^*|0) \in (0,1)$. *Thus, if, for some* u, $z \mapsto h(z,u)$ *is not constant, then* $\pi_1(\cdot)$ *cannot be constant. If* Z *is continuous and* $P(z)$ *includes nonlinear terms,* (3.2) *may hold as well. For instance,* (3.2) *holds with* $P(z) = (1, z, ..., z^k)$, $k \geq 2$, *provided that for some* x, *the polynomial* $z \mapsto P(z)'\pi(x)$ *is not one-to-one. This automatically holds if this polynomial has an even degree, for instance.*

The following theorem shows that, under Assumptions 1 and 4, we can jointly test the exogeneity of the instrument (Assumption 2), the monotonicity of the first stage (Assumption 3) and the exclusion restriction $g(X,Z,\varepsilon) = g(X,\varepsilon)$.

Theorem 1. *Suppose that Assumptions 1-4 hold. Then* $g(X,Z,\varepsilon) = g(X,\varepsilon)$ *implies*

$$Y|X = x^*, Z = z \sim Y|X = x^*, Z = z', \tag{3.3}$$

where the tuple (x^*, z, z') *is defined in Assumption 4.*

A proof is provided in Appendix A.1. This theorem shows that if the assumptions underlying the control function approach and the local irrelevance condition hold, then we can test the exclusion restriction $g(X,Z,\varepsilon) = g(X,\varepsilon)$ by testing the condition (3.3). Alternatively, Theorem 1 may be seen as a way of testing jointly the control function approach (namely, Assumptions 2-3) and the exclusion restriction $g(X,Z,\varepsilon) = g(X,\varepsilon)$.

Note that (3.3) only tests an implication of the exclusion restriction, namely that $g(x^*,z,\varepsilon) = g(x^*,z',\varepsilon)$ for (x^*,z,z') as in Assumption 4. Hence, (3.3) may hold even if $z \mapsto g(x,z,\varepsilon)$ is not constant for some $x \neq x^*$, if the effect of Z on Y is heterogeneous. An example of such a model is

$$\begin{cases} Y &= X\alpha_0 + XZ\beta_0 + \varepsilon, \\ X &= Z\eta, \end{cases}$$

with $\beta_0 \neq 0$ and where $\text{Supp}(Z) \subset (0,\infty)$ and $[-a,a] \subset \text{Supp}(\eta)$ for some $a > 0$. For any $(z,z') \in \text{Supp}(Z)^2$, $z \neq z'$, $F_{X|Z}(\cdot|z)$ and $F_{X|Z}(\cdot|z')$ cross at $x^* = 0$ and at this point, $F_{Y|X=x^*,Z=z} = F_{\varepsilon|\eta=0}$. Thus, (3.3) holds even though the exclusion restriction fails. More generally, it seems however unlikely that Z has no direct effect on Y precisely for the subpopulation $X = x^*$.

3.2 A statistical test

Consider for simplicity that Y is continuous and Z is binary with support $\mathcal{Z} = \{0,1\}$.[6] Suppose that there is a unique x^* such that $F_{X|Z}(x^*|0) = F_{X|Z}(x^*|1) \in (0,1)$, as in Assumption 4 adapted the current setting. Our null hypothesis, based on the testable implication (3.3), is then

$$H_0 : F_{Y|X,Z}(\,\cdot\,|x^*, 0) = F_{Y|X,Z}(\,\cdot\,|x^*, 1).$$

Suppose that we have a random sample $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ of size n. Our proposed testing procedure is based on an estimator for $F_{Y|X,Z}(y|x^*, z)$, i.e.,

$$\widehat{F}_{Y|X,Z}(y|\widehat{x}^*, z) = \frac{\sum_{i=1}^n \mathbb{1}\{Y_i \le y\} \cdot K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}}{\sum_{i=1}^n K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}},$$

where $K_{h_n} = K(\,\cdot\,/h_n)$ for a kernel function K and a bandwidth parameter h_n. The estimator $\widehat{x}^*$ of x^* is defined by

$$\widehat{x}^* \in \arg \min_{x \in \left[\widehat{F}^{-1}_{X|Z}(\underline{p}|0), \widehat{F}^{-1}_{X|Z}(\overline{p}|0)\right]} \left| \widehat{F}_{X|Z}(x|0) - \widehat{F}_{X|Z}(x|1) \right|, \tag{3.4}$$

where $0 < \underline{p} < \overline{p} < 1$ are two constants used to avoid that $\widehat{x}$ tends to x such that $F_{X|Z}(x|0) = F_{X|Z}(x|1) \in \{0,1\}$ and the estimator $\widehat{F}_{X|Z}(\cdot|z)$ is defined by

$$\widehat{F}_{X|Z}(x|z) = \frac{\sum_{i=1}^n \mathbb{1}\{X_i \le x\} \cdot \mathbb{1}\{Z_i = z\}}{\sum_{i=1}^n \mathbb{1}\{Z_i = z\}}.$$

Let $\mathcal{S}_{XYZ} = \text{Supp}(X,Y,Z)$, $\mathcal{S}_{XZ} = \text{Supp}(X,Z)$ and $\mathcal{S}_{YZ} = \text{Supp}(Y,Z)$. Our asymptotic result below relies on the following assumption.

Assumption 5. *(i) There exists a unique x^* satisfying $F_{X|Z}(x^*|0) = F_{X|Z}(x^*|1) \in (0,1)$. Moreover, $F_{X|Z}(x^*|0) \in (\underline{p}, \overline{p})$, with $(\underline{p}, \overline{p})$ introduced above.*

(ii) The conditional density function $f_{X|Z}$ exists and for all $z \in \mathcal{Z} = \{0,1\}$, $f_{X|Z}(\cdot|z)$ is Lipschitz continuous. Moreover, $f_{X|Z}(x^*|0) \neq f_{X|Z}(x^*|1)$ and $f_{X|Z}(x^*|0) \wedge f_{X|Z}(x^*|1) > 0$.

(iii) For all $(x,z) \in \mathcal{S}_{XZ}$, $y \mapsto F_{Y|X,Z}(y|x,z)$ is Lipschitz. For all $(y,z) \in \mathcal{S}_{YZ}$, $x \mapsto F_{Y|X,Z}(y|x,z)$ is

[6] Our procedure directly extends to a finitely supported Z, with just one additional complication from which we abstract here: there may be pairs (z, z') for which $F_{X|Z}(\cdot|z)$ and $F_{X|Z}(\cdot|z')$ do not cross.

continuously differentiable, with $(x,y,z) \mapsto \partial_x F_{Y|X,Z}(y|x,z)$ *bounded on* $\mathcal{S}_{XYZ}$ *and* $x \mapsto \partial_x F_{Y|X,Z}(y|x,z)$ *Lipschitz continuous, with uniform Lipschitz constant* L_F *over* $\mathcal{S}_{YZ}$*.*

(iv) $nh_n^2 \to \infty$ *and* $nh_n^3 \to 0$ *as* $n \to \infty$*.*

(v) $K : [-1,1] \to \mathbb{R}$ *is Lipschitz continuous with constant* L_K*.*

Apart from (i), which is specific to our context, Assumption 5 imposes a set of conditions that are standard in nonparametric settings. In part (iv), we require a certain range of undersmoothing rates of bandwidths. There is in general no data-driven rule to choose an undersmoothing sequence of tuning parameters. We therefore propose the following rule of thumb. Suppose that $\hat{h}_n^*$ is some data-driven optimal bandwidth, which can be obtained by either a plug-in AMISE optimal or cross validation. To adjust this optimal bandwidth into the admissible rate of $n^{-5/12}$ for instance while keeping the constant, we use $\hat{h}_n^* \cdot n^{1/5-5/12}$ as a rule of thumb.

Using in particular Lemma 3 in Appendix B, we establish that under Assumption 5,

$$\sqrt{nh_n}(\widehat{F}_{Y|X,Z}(y|\widehat{x}^*,z) - F_{Y|X,Z}(y|x^*,z)) = \nu_n(y,x^*,z) + o_P(1)$$

uniformly over (y,z), where the uniform influence function representation takes the form of

$$\nu_n(y,x^*,z) = \sum_{i=1}^n \frac{(\mathbb{1}\{Y_i \le y\} - F_{Y|X,Z}(y|X_i,z)) \cdot K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}}{\sqrt{nh_n} f_{X|Z}(x^*|z) \cdot Pr(Z=z)}.$$

Furthermore, its limit Gaussian process $\mathbb{G}$ can be approximated by the multiplier process

$$\nu_{\zeta,n}(y,x^*,z) = \sum_{i=1}^n \zeta_i \frac{(\mathbb{1}\{Y_i \le y\} - F_{Y|X,Z}(y|X_i,z)) \cdot K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}}{\sqrt{nh_n} f_{X|Z}(x^*|z) \cdot Pr(Z=z)},$$

where $\{\zeta_i\}_{i=1}^n$ are independent standard normal variables, independent of the data $\{(Y_i,X_i,Z_i)\}_{i=1}^n$. This multiplier process is infeasible to simulate, though, because we do not know $F_{Y|X,Z}$, x^*, $f_{X|Z}$ or $\Pr(Z=\cdot)$. Therefore, we approximate this multiplier process by the estimated multiplier process

$$\widehat{\nu}_{\zeta,n}(y,\widehat{x}^*,z) = \sum_{i=1}^n \zeta_i \frac{(\mathbb{1}\{Y_i \le y\} - \widehat{F}_{Y|X,Z}(y|X_i,z)) \cdot K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}}{\sqrt{nh_n} \widehat{f}_{X|Z}(\widehat{x}^*|z) \cdot \widehat{\Pr}(Z=z)},$$

where $\widehat{\Pr}(Z=z) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{Z_i = z\}$, $\widehat{f}_{X|Z}(x^*|z) = \frac{\frac{1}{nh_n}\sum_{i=1}^n K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}}{\widehat{\Pr}(Z=z)}$ and $\widehat{x}^*$ is given in (3.4).

Based on the multiplier bootstrap with this estimated multiplier process, we propose the following procedure to test the exclusion restriction. First, form the Kolmogorov-Smirnov-type statistic

$$\widehat{T}_n = \sqrt{nh_n} \left\| \widehat{F}_{Y|X,Z}(\cdot \,|\widehat{x}^*, 0) - \widehat{F}_{Y|X,Z}(\cdot \,|\widehat{x}^*, 1) \right\|_\infty ,$$

where $\|\cdot\|_\infty$ denotes the uniform norm. Second, let $\widehat{c}_n(1-\alpha)$ denote the $(1-\alpha)$-quantile of the conditional distribution of $\|\widehat{\nu}_{\zeta,n}(\cdot \,, \widehat{x}^*, 0) - \widehat{\nu}_{\zeta,n}(\cdot \,, \widehat{x}^*, 1)\|_\infty$ given the data. Our proposed test rejects the null hypothesis H_0 if $\widehat{T}_n > \widehat{c}_n(1-\alpha)$. The following theorem presents the asymptotic level and power properties of this testing procedure.

Theorem 2. *Suppose that Assumption 5 holds and fix $\alpha \in (0,1)$:*

1. *If H_0 holds,* $\Pr\left(\widehat{T}_n > \widehat{c}_n(1-\alpha)\right) \to \alpha$ *as $n \to \infty$;*

2. *If H_0 does not hold,* $\Pr\left(\widehat{T}_n > \widehat{c}_n(1-\tau)\right) \to 1$ *as $n \to \infty$.*

A proof is provided in Appendix A.2. The use of a Kolmogorov-Smirnov-type statistic ensures that the test has power against fixed alternatives of the form $F_{Y|X,Z}(y|x^*, 0) \neq F_{Y|X,Z}(y|x^*, 1)$ for some y. Although we focus on this type of statistic for the sake of conciseness, we remark that similar results follow immediately from the proof for tests based on other (e.g., Cramer-von-Mises-type) statistics.

We do not impose support restrictions on Y to obtain Theorem 2. As for the usual empirical cdf, the properly normalized kernel estimator of the conditional cdf converges uniformly towards a Gaussian process without any such restriction. This is also the case of the multiplier process $y \mapsto \nu_{\zeta,n}(y, x^*, z)$ and its feasible version $y \mapsto \widehat{\nu}_{\zeta,n}(y, \widehat{x}^*, z)$. We refer to the proof of Theorem 2 and Lemmas 3–5 in Appendix B for more details, and to Horvath and Yandell (1988) for an early result of this kind.

3.3 A Monte Carlo Simulation Study

In this section, we present a Monte Carlo simulation study to examine the finite sample performance of the proposed statistical test of the exclusion restriction.

Consider the nonseparable model with a possibly included instrument:

$$\text{DGP 1:} \qquad Y = \varepsilon_0 + \varepsilon_1 X + \beta Z$$
$$X = \eta \cdot (1 + Z).$$

The binary instrument $Z \sim \text{Bernoulli}(0.5)$ is generated independently of the trivariate unobservables $(\eta, \varepsilon_0, \varepsilon_1) \sim \mathcal{N}(0, E_3/4 + 3I_3/4)$, where E_p (resp. I_p) is the matrix of ones (resp. the identity matrix) of size p. The exclusion restriction holds if $\beta = 0$ and is violated otherwise. The outcome equation is a particular case of the linear random coefficient model considered in Section 4.1 below.

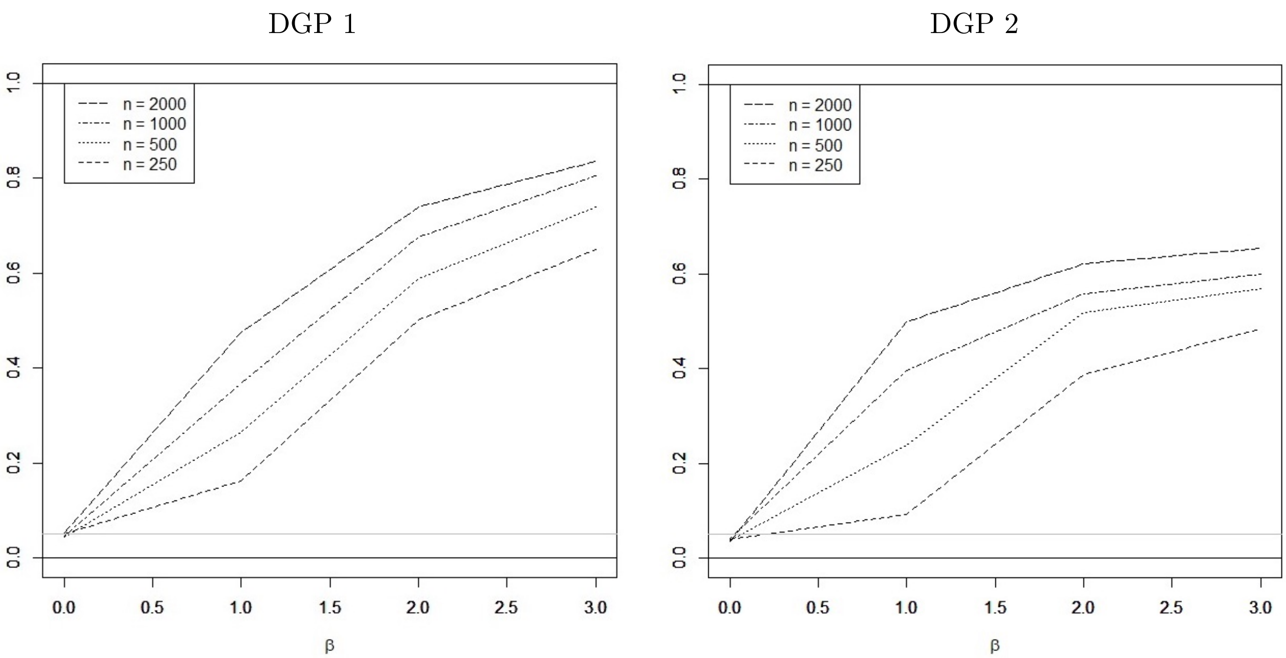
We also consider the following binary choice model:

$$\text{DGP 2:} \qquad Y = \mathbb{1}\left\{X + \beta Z + \varepsilon \geq 0\right\},$$
$$X = \eta \cdot (1 + Z),$$

where Z is as above and the bivariate unobservables are generated according to $(\eta, \varepsilon) \sim \mathcal{N}(0, E_2/4 + 3I_2/4)$. The outcome equation is thus a probit model with an endogenous covariate (X) and the exclusion restriction holds if $\beta = 0$ and is violated otherwise. The model satisfies the index restriction considered in Section 4.2.

We consider tests with a nominal size of 5%. For both DGPs, we vary the sample size and the value of β across sets of Monte Carlo simulations. In each set, we use 2,500 samples to compute the frequencies of acceptance/rejection by our test. For each such sample, we use 2,500 multiplier bootstrap iterations to compute the critical value of the test. Following the guidance presented in Section 3.2, we use $\hat{h}_n^* \cdot n^{1/5-5/12}$ as the bandwidth parameter, where $\hat{h}_n^*$ is the leave-one-out cross-validation optimal choice with squared error loss.

Figure 1 depicts the power curves for DGP 1 (left) and DGP 2 (right). At $\beta = 0$, the rejection frequencies are approximately the same as the nominal size, 0.05, even with $n = 250$, for both DGPs. This suggests two conclusions. First, our bandwidth choice implies a small bias, as needed for our theory. Second, the multiplier bootstrap appears to approximate well the distribution of the test statistic under the null, even if n is relatively small. Figure 1 also shows that as one could expect, the rejection frequencies increase with the sample size and the value of β. To give a sense of the power, note that for DGP1 and when $\beta = 1$, βZ contributes no more than 6.2% of the total variance of Y. Still, for such a β, the power is already equal to 26.4% with $n = 500$, and 47.6% with $n = 2,000$. A similar conclusion holds for DGP2, though the power appears to be lower in this case.



Notes: β indexes the extent of instrument inclusion. The nominal size of the test is 0.05.

Figure 1: Power curves of the test.

4 Identification without exclusion restriction

Theorem 1 suggests that, under Assumptions 1-4, the exclusion restriction $g(X, Z, \varepsilon) = g(X, \varepsilon)$ may not be necessary for the identification of causal effects. We provide results in this direction in this section, under other restrictions on g. Such results may be useful in particular if we reject the previous test, but still Assumptions 1-3 appear credible.[7] For simplicity, we refer to Z as the instrument hereafter. One must keep in mind, however, that it may not satisfy the exclusion restriction, which

[7]If, instead, the exclusion restriction is credible (as in, e.g., randomized experiments), one can rather relax the first-stage monotonicity by following the approach of Masten (2017) or Hoderlein et al. (2017).

means that it may have a direct effect on Y.

4.1 Linear random coefficients models

The first leading case we consider is linear random coefficients models, which have been extensively studied in the literature, see, e.g., Beran et al. (1992); Hoderlein et al. (2010). The specification in our notation is given by the following assumption.

Assumption 6 (Linear Random Coefficient Model). $g(X, Z, \varepsilon) = \varepsilon_0 + \varepsilon_1 X + \varepsilon_2 Z$.

In a control function setup, Florens et al. (2008) and Masten and Torgovitsky (2016) study a similar model to the random coefficients model in Assumption 6, but under the key exclusion restriction that $\varepsilon_2 = 0$. Thus, our model allows not only for heterogeneous treatment effects potentially correlated with the treatment itself, but also for a direct, possibly heterogeneous, effect of the instrument Z. The following example, adapted from Florens et al. (2008), shows that Assumptions 1-4 and 6 may be derived from a structural choice model on X.

Example 3. *Assume that X is the level of schooling, Y denotes observed wage (discounted annualized earning flows) and Z corresponds to, e.g., college subsidies. As Florens et al. (2008), assume that the potential wages (discounted annualized earning flows) $Y(x)$ satisfy $Y(x) = \varepsilon_0 + \varepsilon_1 x + \varepsilon_2 Z$. We assume that the cost of schooling satisfies*

$$C(x) = C_0(Z) + (C_1(Z) + \nu_1)x + \frac{C_2(Z)}{2}x^2 + \nu_0.$$

We thus make two differences compared to Florens et al. (2008). First, we rule out nonlinear effects of schooling, while they include an additional term $\varphi_2 x^2/2$ (with φ_2 constant). Second, following Jones (2015), we allow for a direct effect of Z on earnings. If individuals choose their education level to maximize wages minus cost, we obtain $X = (-C_1(Z) + \varepsilon_1 - \nu_1)/C_2(Z)$. If $(\varepsilon, \nu_0, \nu_1)$ are independent of Z, Assumptions 2 and 3 hold, with $\eta = F_{\varepsilon_1 - \nu_1}(\varepsilon_1 - \nu_1)$. Since the first stage is a location-scale model, Assumption 4 holds if either $C_2(\cdot)$ is not constant or there exist $z_0 \neq z_1$ such that $C_1(z_0) = C_1(z_1)$.

We first consider the identification of $E(\varepsilon_1|\eta = u)$ and $E(\varepsilon_2|\eta = u)$, the average marginal effect (or discrete change if Z is binary) of X and Z for the subpopulation for which $\eta = u$. Assumptions 1-4 and 6 are sufficient to achieve identification of $E(\varepsilon_1|\eta = u)$ and $E(\varepsilon_2|\eta = u)$ for suitable u. Specifically, we show in the proof of Theorem 3 below that $E(\varepsilon_2|\eta = u)$ is identified for all $u \in \mathscr{C}$, with

$$\mathscr{C} = \left\{u \in (0,1) : \exists (x, z, z'), (z, z') \in \text{Supp}(Z)^2, z' \neq z : F_{X|Z}(x|z) = F_{X|Z}(x|z') = u\right\}.$$

To see this, let (x^*, z, z') be as in Assumption 4. The set $\mathscr{C}$ is not empty since it includes $u^* = F_{X|Z}(x^*|z)$. Now, focusing here on these elements (x^*, z, z'), when Z moves from z to z' while X is kept constant and equal to x^*, η also remains constant and equal to u^*. Thus, any change in Y must come from the direct effect of Z, meaning that we identify $E(\varepsilon_2|\eta = u^*)$.

Theorem 3 below also shows that $E(\varepsilon_1|\eta = u)$ is identified for all $u \in \mathscr{C}'$, with

$$\mathscr{C}' = \left\{u \in \mathscr{C}' : \exists (x, z, z') : u = F_{X|Z}(x|z) \neq F_{X|Z}(x|z')\right\}.$$

To see this, let $\widetilde{Y} = Y - E(\varepsilon_2|\eta)Z$. Then a move of Z that affects X, while keeping η constant, will impact the average of $\widetilde{Y}$ only through X. This allows us to recover $E(\varepsilon_1|\eta = u)$ for $u \in \mathscr{C}'$. However, when $Z \in \{0, 1\}$, we have $\mathscr{C}' = \emptyset$: if $u \in \mathscr{C}$, $F_{X|Z}(x|0) = F_{X|Z}(x|1)$, which implies that $u \notin \mathscr{C}'$.

Hence, under Assumptions 1-4 and 6 alone, marginal effects are identified for some subpopulations. We can then identify the average effect on the whole population $E(\varepsilon_2) = E(E(\varepsilon_2|\eta))$ if $\mathscr{C} = (0, 1)$. Similarly, we identify $E(\varepsilon_1)$ if $\mathscr{C}' = (0, 1)$. These two conditions hold, for instance, if there exist $z \neq z'$ and $z(\cdot)$ such that $F_{X|Z}(\cdot|z) = F_{X|Z}(\cdot|z')$ and for all x, $F_{X|Z}(x|z) \neq F_{X|Z}(x|z(x))$. In cases where $\mathscr{C} \neq (0, 1)$ or $\mathscr{C}' \neq (0, 1)$, we can still identify $E(\varepsilon_1)$ and $E(\varepsilon_2)$ under the following additional restrictions.

Assumption 7 (Exogenous random effect of Z)**.** $E(\varepsilon_2|\eta) = E(\varepsilon_2)$.

Assumption 8 (Global Relevance of the Instrument)**.** *There exists* $(z_1, z_2) \in \mathcal{Z}^2$ *such that* $F_{X|Z}(X|z_1) \neq F_{X|Z}(X|z_2)$ *almost surely.*

Assumption 7, when combined with Assumption 2, implies that $E(\varepsilon_2|X) = E(\varepsilon_2)$. In this sense, ε_2 is exogenous, contrary to $(\varepsilon_0, \varepsilon_1)$. A particular case where this holds is when ε_2 is actually constant, as in standard linear models. But it may also hold if common factors affect both ε_2 and $(\varepsilon_0, \varepsilon_1)$, provided that these factors are independent of (Z, η). In Assumption 8, (z_1, z_2) may actually be equal to the pair (z, z') appearing in Assumption 4. If $F_{X|Z}(\cdot|z)$ and $F_{X|Z}(\cdot|z')$ cross once, or even a countable number of times, then Assumption 8 holds with $(z_1, z_2) = (z, z')$. Thus, Assumptions 4 and 8 can simultaneously hold, even if the instrument is binary. This occurs typically in a large class of generalized location-scale models.

Example 1 (Continued)**.** *Assumption 8 holds if and only if either $\psi(\cdot)$ or $\sigma(\cdot)$ (or both) is non-constant – see Appendix C.2 for a proof. Thus, if either $\psi(\cdot)$ is not constant and not one-to-one, or $\sigma(\cdot)$ is not constant, then both Assumptions 4 and 8 are satisfied.*

As a result, in the random coefficients model above, $E(\varepsilon_1)$ and $E(\varepsilon_2)$ can be identified even with a binary instrument. We can actually identify the entire model under additional, more stringent conditions on the instrument. To state those, we need the notation $\Psi_{A|B}$ to denote the conditional characteristic function of a random variable A given a random variable B.

Assumption 9 (Independent random effect of Z)**.** *ε_2 is independent of $(\varepsilon_0, \varepsilon_1, \eta)$.*

Assumption 10. *(i) For all $u \in (0, 1)$, there exists a point x_u and a sequence $(x_{n,u})_{n\geq 1}$, both in Supp$(h(Z, u))$, such that $x_{n,u} \to x_u$ and for all n, $x_{n,u} \neq x_u$.*
(ii) Ψ_{ε_2} and $\Psi_{\varepsilon_0+\varepsilon_1 x^*|\eta=F_{X|Z}(x^*|z)}$, with (x^*, z) defined in Assumption 4, do not vanish on the real line.
(iii) $E(\exp(c|\varepsilon_1|)) < +\infty$ for some $c > 0$.

Assumption 9 reinforces Assumption 7, in particular, by ruling out any dependence between ε_2 and $(\varepsilon_0, \varepsilon_1)$. Assumption 10 (i) implies that Z takes an infinite number of values and has an effect on X, which may be interpreted as a relevance condition on Z. Note, on the other hand, that Condition (i) does not require a large support of Z. It holds, for instance, if Z is continuous and $z \mapsto h(z, u)$ is continuous and non-constant. Condition (ii) is a non-vanishing condition that is standard in deconvolution problems. Condition (iii) requires that the tails of the marginal effect ε_1

are thin enough. This condition could be avoided, but at the expense of a large support requirement on $h(Z, u)$, which is rarely satisfied in practice.

The following theorem shows the identification of the local parameters $E(\varepsilon_1|\eta = u), E(\varepsilon_1|\eta = u)$, the global parameters $E(\varepsilon_1)$ and $E(\varepsilon_2)$, and the entire model under combinations of the previous assumptions.

Theorem 3. *Suppose that Assumptions 1–4 and 6 hold. Then:*

1. *$E(\varepsilon_1|\eta = u)$ (resp. $E(\varepsilon_2|\eta = u)$) is identified for all $u \in \mathscr{C}'$ (resp. $u \in \mathscr{C}$). $E(\varepsilon_1)$ (resp. $E(\varepsilon_2)$) is then identified if $\mathscr{C}' = (0, 1)$ (resp. $\mathscr{C} = (0, 1)$).*

2. *If Assumptions 7-8 also hold, then $E(\varepsilon_1)$, $E(\varepsilon_2)$, $E(\varepsilon_1|\eta = u)$ and $E(\varepsilon_2|\eta = u)$ are identified for all $u \in (0, 1)$.*

3. *If Assumptions 9-10 also hold, the distribution of $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \eta)$ is identified.*

The third part implies in particular that all the average treatment effects $E(\varepsilon_1|\eta = u)$ and $E(\varepsilon_2|\eta = u)$ are identified, as in the second part. But it also implies that any quantile treatment effects of X and Z, and the distribution of the individual causal effects (i.e., the distribution of ε_1) are identified.

4.2 Index on X and Z

Consider the following index restriction on the structural function.

Assumption 11 (Index Restriction on (X, Z))**.** *(i) $g(X, Z, \varepsilon) = g_1(x\alpha_0 + z\beta_0, \varepsilon)$ with $\beta_0 \in \{-1, 0, 1\}$ and $\alpha_0 \in \{-1, 0, 1\}$ if $\beta_0 = 0$;*
(ii) For all $u \in (0, 1)$, the map $w \mapsto E(g_1(w, \varepsilon)|\eta = u)$ is strictly increasing on Supp$(X\alpha_0 + Z\beta_0|\eta = u)$.

Whereas Assumption 11 is compatible with non-additive errors, contrary to the random coefficients model above, the index restriction imposes restrictions on the heterogeneity of effects. In particular, it implies that the relative effect $\partial_x g(x, z, e)/\partial_z g(x, z, e)$, assuming here that the derivatives exist, is a constant function of (x, z, e). We shall see below that there are testable implications of this restriction. The condition $\beta_0 \in \{-1, 0, 1\}$ is a mere normalization: if $\beta_0 \neq 0$ and $|\beta_0| \neq 1$, we can always divide $x\alpha_0 + z\beta_0$ by $|\beta_0|$ and change g_1 accordingly. When $\beta_0 = 0$, this normalization has to be done on α_0 instead. Condition (ii) can be interpreted as an average monotonicity restriction on g, where we take the average over ε, conditional on η.[8] Assumption 11 applies to the linear model $g(x, z, \varepsilon) = x\alpha_0 + z\beta_0 + \varepsilon$ in particular, but also to a wide range of limited dependent variable models. It holds for instance in the binary choice model

$$Y = \mathbb{1}\{X\alpha_0 + Z\beta_0 - \varepsilon \geq 0\}, \tag{4.1}$$

provided that $F_{\varepsilon|\eta}$ is strictly increasing. A classical example of such a model is female participation to the labor market (see in particular Blundell and Powell, 2004), where the endogenous variable X corresponds to other income (e.g., husband's earnings) and Z could be (following, again Blundell and Powell, 2004) the husband's education. This latter variable could have a direct effect on participation through, e.g. the husband's acquaintances.

We first investigate the identification of (α_0, β_0), before turning to average marginal effects. For our results, we also invoke Assumption 12 below. For any $(u, z, z') \in (0, 1) \times \mathcal{Z}^2$, let $F^{-1}_{X|Z}(u|z) = \inf\{x : F_{X|Z}(x|z) \geq u\}$ and $q_{zz'}(x) = F^{-1}_{X|Z}\left[F_{X|Z}(x|z)|z'\right]$. $q_{zz'}(\cdot)$ is thus the quantile-quantile transform between the two distributions $F_{X|Z}(.|z)$ and $F_{X|Z}(.|z')$.

Assumption 12. *There exists $(z_1, z_1') \in \mathcal{Z}^2$ and $x_1 \in Supp(X|Z = z_1)$ such that $q_{z_1 z_1'}(x_1) \neq x_1$. Moreover, we either have:*

(i) $E\left(Y|X = x^*, Z = z\right) = E\left(Y|X = x^*, Z = z'\right)$, with (x^*, z, z') defined in Assumption 4;

or (ii) $E(Y|X = x_1, Z = z_1) = E(Y|X = q_{z_1 z_1'}(x_1), Z = z_1')$.

Assumption 12 only involves observed variables and is therefore testable. The first condition is a relevance condition on the instrument, which is weaker than Assumption 8 above. Then we either impose (i) or (ii). Condition (i) corresponds to the case where, when restricting to the values (x^*, z, z') such that Z has no effect on X, a shift on Z has no effect on $E(Y|X = x^*, Z = z)$. As shown in the proof of Theorem 4 below, this occurs when Z has no direct effect on Y, namely when $\beta_0 = 0$.

[8] Condition (ii) holds automatically when $\alpha_0 = \beta_0 = 0$, since then $\text{Supp}(X\alpha_0 + Z\beta_0|\eta) = \{0\}$.

Alternatively, Condition (ii) states that it is possible to produce compensating variations. To understand that, remark that a shift in Z moves Y directly but also indirectly, through the corresponding change in X, i.e., the shift from x_1 to $q_{z_1 z_1'}(x_1)$.[9] Condition (ii) requires that both effects offset each other for some (x_1, z_1, z_1'). Roughly speaking, this implies that the direct and indirect effects do not always add up. If $\beta_0 > 0$, say, so that the direct effect of Z is positive, Condition (ii) implies that there are values $z_1 < z_1'$ such that a shift from z_1 to z_1' has a negative indirect effect. This means that either $\alpha_0 > 0$ (positive effect of X on Y) and $q_{z_1 z_1'}(x_1) < x_1$, or $\alpha_0 < 0$ and $q_{z_1 z_1'}(x_1) > x_1$. In all cases, $\alpha_0 \neq 0$, namely X has an effect on Y. In the example of female labor participation, Assumption 12 (ii) may be satisfied since husband's education has a negative indirect effect, through the increase of other income, but plausibly a positive, direct effect.

Finally, as in Assumption 8 above, we may actually have $(z, z') = (z_1, z_1')$, with (z, z') defined in Assumption 4. Hence, Assumptions 4 and 12 may both hold with a binary instrument. Similarly to the combination of Assumptions 4 and 8, these two assumptions are satisfied in a location-scale specification of the first-stage equation, as long as X has an effect on Y and there is some heteroskedasticity.

Example 1 (Continued)**.** *Suppose that Assumption 11 holds, $\alpha_0 \neq 0$, $\mu(x) = x$ and $\sigma(\cdot)$ is not constant. Then both Assumptions 4 and 12 hold – see Appendix C.3 for a proof.*

The following theorem shows that under the previous assumptions, (α_0, β_0) is identified.

Theorem 4. *Suppose that Assumptions 1-4 and 11-12 hold. Then, (α_0, β_0) is identified.*

The proof can be summarized as follows. First, letting (x^*, z, z') be as in Assumption 4, we establish that the sign of $E(Y|X = x^*, Z = z') - E(Y|X = x^*, Z = z)$ identifies β_0. Second, focusing here on the case $\beta_0 \neq 0$, we show that α_0 is identifed by the compensating variation condition in Assumption 12(ii). Specifically, letting (x_1, z_1, z_1') be as in Assumption 12, we show that

$$\alpha_0 = \frac{(z_1 - z_1')\beta_0}{q_{z_1 z_1'}(x_1) - x_1}.$$

[9] Since the change in X results from an exogenous change in Z, there is no other effect due to a change in the distribution of ε.

This equation also implies that, if we observe other points (x_2, z_2, z_2') satisfying Assumption 12, then we can actually test the index restriction, as it implies

$$(z_1 - z_1')(q_{z_2 z_2'}(x_2) - x_2) = (z_2 - z_2')(q_{z_1 z_1'}(x_1) - x_1).$$

We now turn to the identification of average marginal effects, supposing here that Z is continuous. Specifically, we impose the following condition, which involves the index $W = X\alpha_0 + Z\beta_0$.

Assumption 13. *(i) For almost all $u \in (0,1)$, Supp$(W|\eta = u)$ is a non-trivial interval.*
(ii) for almost all $u \in (0,1)$, $w \mapsto E[g_1(w, \varepsilon)|\eta = u]$ is differentiable.

Under Assumptions 2-3, $\text{Supp}(W|\eta = u) = \text{Supp}(h(Z, u)\alpha_0 + Z\beta_0)$. Thus, Condition (i) holds if $h(\cdot, u)$ is continuous, $\mathcal{Z}$ is a non-degenerate interval and either $\beta_0 \neq 0$ or $\alpha_0 \neq 0$ and $h(\cdot, u)$ is not constant. As is the case with the monotonicity condition in Assumption 11, Condition (ii) may hold even if Y is a limited dependent variable. If Y is binary and (4.1) holds, for instance, then Condition (ii) is satisfied as long as $F_{\varepsilon|\eta}(\cdot|u)$ is differentiable for almost all u.

To define the average marginal effects we consider, let us introduce

$$m(x_1, x_2, z_1, z_2) = E\left[g(x_1, z_1, \varepsilon)|X = x_2, Z = z_2\right], \tag{4.2}$$

the average counterfactual outcome for units such that $(X, Z) = (x_2, z_2)$, if their (X, Z) was set to (x_1, z_1). Then we consider the local average marginal effects

$$\Delta_X(x, z) = \partial_{x_1} m(x, x, z, z),$$
$$\Delta_Z(x, z) = \partial_{z_1} m(x, x, z, z).$$

Note that the derivatives are well-defined under Assumption 13-(ii). We also consider the global average marginal effects $\Delta_X = E\left[\Delta_X(X, Z)\right]$ and $\Delta_Z = E\left[\Delta_Z(X, Z)\right]$.

Theorem 5. *Suppose that Assumptions 1-4 and 11-13 hold. Then, for all $(x, z) \in$ Supp(X, Z), $\Delta_X(x, z)$, $\Delta_Z(x, z)$, Δ_X and Δ_Z are identified.*

The results for Δ_X and Δ_Z simply follow from $\Delta_X = \alpha_0 E[\partial_w e(W,\eta)]$ and $\Delta_Z = \beta_0 E[\partial_w e(W,\eta)]$, with $e(w,u) = E(Y|W = w, \eta = u)$. As with the random coefficients model, we are thus able to identify more parameters when moving from a discrete to a continuous instrument. In particular, discrete, exogenous change in X induced by a discrete Z are not sufficient in general to point identify Δ_X.

4.3 Index on X and ε

We now consider another index restriction, this time on X and ε. The advantage of this approach is that the index can remain nonparametric. On the other hand, this approach is limited to continuous outcome variables. Specifically, we make the following two assumptions.

Assumption 14 (Continuity and Regularity of Y)**.** *For all $(x,z) \in Supp(X,Z)$, $Supp(Y|X = x, Z = z)$ is an interval, which does not depend on x.*

Assumption 15 (Index Restriction on (X,ε))**.** *(i) $g(X,Z,\varepsilon) = g_1(Z,\varphi(X,\varepsilon))$, where $\varphi(X,\varepsilon) \in \mathbb{R}$ and $g_1(z,\cdot)$ is strictly increasing for all $z \in \mathcal{Z}$.*
(ii) There exists $z_1 \in \mathcal{Z}$ such that $g_1(z_1,u) = u$ for all $u \in Supp(Y|Z = z_1)$.
(iii) $Supp(\varphi(X,\varepsilon)|Z = z)$ does not depend on z and $g_1(z,\cdot)$ is continuous at the boundary of $Supp(\varphi(X,\varepsilon))$.

Assumption 14 rules out discrete dependent variables. Assumption 15 (i) imposes an index restriction between X and ε in g. Contrary to the index on X and Z imposed above, we do not restrict this index to be linear. Condition (ii) is a mere normalization, since we can always replace $\varphi(X,\varepsilon)$ and $g_1(z_1,\cdot)$ by $g_1(z_1,\varphi(X,\varepsilon))$ and the identity function, respectively. Condition (iii) is a mild regularity condition.

In order to identify causal effects in this model, we have to strengthen the local irrelevance condition on the instrument. Hereafter, we say that z and z' are connected if there exists $n \in \mathbb{N}$ and $z_0, ..., z_n$ with $z_0 = z$ and $z_n = z'$ such that for all $k \in \{0, ..., n-1\}$, there exists x_k such that $F_{X|Z}(x_k|z_k) = F_{X|Z}(x_k|z_{k+1}) \in (0,1)$. The local irrelevance condition above ensures that there exist $z \neq z'$ in $\mathcal{Z}$ that are connected. We strengthen this condition by imposing the following assumption.

Assumption 16 (Locally Irrelevant Instrument – Strong Version)**.** *There exists* $z_0 \in \mathcal{Z}$ *such that any* $z \in \mathcal{Z}$ *is connected to* z_0*.*

Assumption 16 is actually equivalent to Assumption 4 when Z is binary, but stronger otherwise. We show below that it holds in the generalized location-scale model (3.1) if, again, there is some heterosckedasticity.

Assumptions 14-16 are compatible with a discrete, even binary, instrument. Theorem 6 below states that these assumptions, together with Assumptions 1-3, are sufficient to identify some discrete treatment effects. More precisely, we recover, for some (x, x', z, z'), $\Delta_X(x, x', z)$ and $\Delta_Z(x, z, z')$ defined by

$$\Delta_X(x, x', z) = m(x', x, z, z) - m(x, x, z, z),$$

$$\Delta_Z(x, z, z') = m(x, x, z', z) - m(x, x, z, z).$$

where m is given in (4.2). Assumptions 14-16, however, are not sufficient to identify marginal effects. We impose for that purpose the following additional condition, which is similar to Assumption 13.

Assumption 17 (Continuous Z and Strong Global Relevance of the Instrument)**.** $\mathcal{Z}$ *is a non-degenerate interval and for almost all* $(z, u) \in \mathcal{Z} \times (0, 1)$ *and all neighborhoods* $\mathcal{V} \subset \mathcal{Z}$ *of* z*,* $h(\cdot, u)$ *is not constant on* $\mathcal{V}$*.*

Importantly, Assumption 17 is compatible with the strong local irrelevance condition, as we illustrate in our example of generalized location-scale models.

Example 1 (Continued)**.** *Suppose that* $\sigma(\cdot)$ *is not constant. Then Assumption 16 holds. Moreover, if* $\mathcal{Z}$ *is a non-degenerate interval and* 1 *(the constant function),* $\psi(\cdot)$ *and* $\sigma(\cdot)$ *are linearly independent on any interval, Assumption 17 holds as well – see Appendix C.4 for proofs.*

Theorem 6. *Suppose that Assumptions 1-3 and 14-16 hold. Then,*

1. *for all* (x, z, z') *such that* $(x, z) \in$ *Supp*(X, Z) *and* $z' \in \mathcal{Z}$*,* $\Delta_X(x, q_{zz'}(x), z)$ *and* $\Delta_Z(x, z, z')$ *are identified;*

2. *if Assumptions 17 and 18 (displayed in Appendix A.6) also hold,* $\Delta_X(x,z), \Delta_Z(x,z), \Delta_X$ *and* Δ_Z *are identified as well, for all* $(x,z) \in Supp(X,Z)$.

A proof is provided in Appendix A.6. The result on $\Delta_X(x,z,z')$ is similar to the identification of the ATT in Theorem 1 of D'Haultfoeuille et al. (2015). In that paper, a similar model is considered but in the context of repeated cross sections, where time plays the role of the instrument Z here. Because time is discrete in that setting, no point identification result on Δ_X or on other marginal effect is given in D'Haultfoeuille et al. (2015), so the second part of Theorem 6 is new. As with the random coefficients model or with index on X and Z, it showcases the benefit of moving from a discrete to a continuous instrument. While in the former case, we can identify some discrete effects of X and Z, in the latter case, we can also identify their average marginal effects on the whole population.

5 Conclusion

In this paper, we first show that under the control function approach, we can test the exclusion restriction in nonseparable triangular models as soon as a local irrelevance condition on the instrument holds, and we devise and analyze such a test. We also show that causal effects can be identified without exclusion restrictions in the linear random coefficients models or under index restrictions. Identification of some effects can be achieved even if the instrument is binary, but other effects require the instrument to have a richer support. We hope that these encouraging theoretical results spurn interest in applications, which we leave for future research.

References

Altonji, J., Elder, T. E. and Taber, C. R. (2005), 'An evaluation of instrumental variable strategies for estimating the effects of catholic schooling', *Journal of Human Resources* **40**, 791–821.

Beran, R., Hall, P. et al. (1992), 'Estimating coefficient distributions in random coefficient regressions', *nnals of Statistics* **20**(4), 1970–1984.

Blundell, R. W. and Powell, J. L. (2004), 'Endogeneity in semiparametric binary response models', *The Review of Economic Studies* **71**(3), 655–679.

Bound, J. and Jaeger, D. A. (2000), Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?, *in* 'Research in Labor Economics', Vol. 19, Bingley, West-Yorkshire: Emerald Group Publishing Limited, pp. 83–108.

Caetano, C., Rothe, C. and Yıldız, N. (2016), 'A discontinuity test for identification in triangular nonseparable models', *Journal of econometrics* **193**(1), 113–122.

Chesher, A. (2003), 'Identification in nonseparable models', *Econometrica* **71**, 1405–1441.

D'Haultfoeuille, X. and Février, P. (2015), 'Identification of nonseparable models with endogeneity and discrete instruments', *Econometrica* **83**, 1199–1210.

D'Haultfoeuille, X., Hoderlein, S. and Sasaki, Y. (2015), Nonlinear difference-in-differences in repeated cross sections with continuous treatments. Working paper.

Feng, J. (2020), Matching points: Supplementing instruments with covariates in triangular models. Working paper.

Florens, J., Heckman, J. J., Meghir, C. and Vytlacil, E. (2008), 'Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects', *Econometrica* **76**, 1191–1206.

Foresi, S. and Peracchi, F. (1995), 'The conditional distribution of excess returns: An empirical analysis', *Journal of the American Statistical Association* **90**(430), 451–466.

Gautier, E. and Hoderlein, S. (2015), A triangular treatment effect model with random coefficients in the selection equation. TSE Working Paper 598.

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica: Journal of the econometric society* pp. 153–161.

Hoderlein, S., Holzmann, H. and Meister, A. (2017), 'The triangular model with random coefficients', *Journal of econometrics* **201**(1), 144–169.

Hoderlein, S., Klemelä, J. and Mammen, E. (2010), 'Analyzing the random coefficient model nonparametrically', *Econometric Theory* **26**(3), 804–837.

Hoderlein, S. and Sasaki, Y. (2013), Outcome conditioned treatment effects. Working Paper.

Horvath, L. and Yandell, B. S. (1988), 'Asymptotics of conditional empirical processes', *Journal of Multivariate Analysis* **26**(2), 184–206.

Imbens, G. W. and Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475.

Imbens, G. W. and Newey, W. K. (2009), 'Identification and estimation of triangular simultaneous equations models without additivity', *Econometrica* **77**, 1481–1512.

Jones, D. (2015), The economics of exclusion restrictions in iv models, Technical report, National Bureau of Economic Research.

Kato, K. (2019), Lecture notes on empirical process theory, Technical report, Technical Report, Cornell University.

Kitagawa, T. (2015), 'A test for instrument validity', *Econometrica* **83**(5), 2043–2063.

Klein, R. and Vella, F. (2010), 'Estimating a class of triangular simultaneous equations models without exclusion restrictions', *Journal of Econometrics* **154**(2), 154–164.

Kosorok, M. R. (2003), 'Bootstraps of sums of independent but not identically distributed stochastic processes', *Journal of Multivariate Analysis* **84**(2), 299–318.

Lewbel, A. (2012), 'Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models', *Journal of Business & Economic Statistics* **30**(1), 67–80.

Masten, M. A. (2017), 'Random coefficients on endogenous variables in simultaneous equations models', *The Review of Economic Studies* **85**(2), 1193–1250.

Masten, M. A. and Torgovitsky, A. (2016), 'Identification of instrumental variable correlated random coefficients models', *Review of Economics and Statistics* **98**(5), 1001–1005.

Newey, W. K. and Stouli, S. (2018), Heterogenous coefficients, discrete instruments, and identification of treatment effects. arXiv preprint arXiv:1811.09837.

Newey, W. and Stouli, S. (2019), 'Control variables, discrete instruments, and identification of structural functions', *Journal of Econometrics* **Forthcoming**.

Rigobon, R. (2003), 'Identification through heteroskedasticity', *Review of Economics and Statistics* **85**(4), 777–792.

Rosenzweig, M. R. and Wolpin, K. I. (2000), 'Natural" natural experiments" in economics', *Journal of Economic Literature* **38**(4), 827–874.

Torgovitsky, A. (2015), 'Identification of nonseparable models using instruments with small support', *Econometrica* **83**, 1185–1197.

van den Berg, G. J. (2007), 'An economic analysis of exclusion restrictions for instrumental variable estimation'.

van Kippersluis, H. and Rietveld, C. A. (2018), 'An evaluation of instrumental variable strategies for estimating the effects of catholic schooling', *Econometrics Journal* **21**, 316–331.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

Vytlacil, E. and Yildiz, N. (2007), 'Dummy endogenous variables in weakly separable models', *Econometrica* **75**, 757–779.

Wooldridge, J. M. (2015), 'Control function methods in applied econometrics', *Journal of Human Resources* **50**(2), 420–445.

A Proofs of the main results

A.1 Theorem 1

We show that if Assumptions 1-4 hold and $g(X,Z,\varepsilon)=g(X,\varepsilon)$, (3.3) is satisfied. By the independence between Z and (ε,η), and because the $\sigma(\cdot)$-algebras generated by (X,Z) and by (Z,η) are the same, we have, for any measurable set A and any $(x,z)\in \text{Supp}(X,Z)$,

$$\begin{aligned}\Pr(\varepsilon\in A|X=x,Z=z)=&\Pr(\varepsilon\in A|Z=z,\eta=h^{-1}(z,x))\\ =&\Pr(\varepsilon\in A|\eta=h^{-1}(z,x)),\end{aligned} \tag{A.1}$$

where $h^{-1}(z,\cdot)$ denotes the inverse of $h(z,\cdot)$. Moreover, using again Assumptions 2-3, we have

$$\Pr(X\leq x|Z=z)=\Pr(h(z,\eta)\leq x)=h^{-1}(z,x). \tag{A.2}$$

Therefore, $F_{X|Z}(x^*|z_1)=F_{X|Z}(x^*|z_2)$ implies that $h^{-1}(z_1,x^*)=h^{-1}(z_2,x^*)$. In view of (A.1), this implies $\varepsilon|X=x^*,Z=z_1\sim\varepsilon|X=x^*,Z=z_2$. Hence,

$$g(x^*,\varepsilon)|X=x^*,Z=z_1\sim g(x^*,\varepsilon)|X=x^*,Z=z_2.$$

The result follows. □

A.2 Proof of Theorem 2

Let us define $\sigma(y,y'^*,z)$ by

$$\sigma(y,y'|x,z)=E\left[\left(\mathbb{1}\{Y\leq y\}-F_{Y|X,Z}(y|X,z)\right)\cdot\left(\mathbb{1}\{Y\leq y'\}-F_{Y|X,Z}(y'|X,z)\right)|X=x,Z=z\right].$$

By Lemmas 3 and 4 in Appendix B.1 and Theorem 1 of Kosorok (2003), we have, under Assumption 5 and for $z\in\{0,1\}$,

$$\sqrt{nh_n}(\widehat{F}_{Y|X,Z}(\,\cdot\,|\widehat{x}^*,z)-F_{Y|X,Z}(\,\cdot\,|x^*,z))\rightsquigarrow\mathbb{G}_z.$$

Here, "$\rightsquigarrow$" denotes weak convergence in the space of bounded functions taking values in $\mathbb{R}$ and $\mathbb{G}_z$ is a zero mean Gaussian process with covariance function

$$(y,y')\mapsto\sigma(y,y'|x^*,z)/\left(f_{X|Z}(x^*|z)\cdot\Pr(Z=z)\right).$$

Next, by Theorem 2 of Kosorok (2003) and under Assumption 5, we have, conditional on the data and for each $z \in \{0, 1\}$,

$$\nu_{\zeta,n}(\,\cdot\,, x^*, z) \rightsquigarrow \mathbb{G}_z.$$

Define the intermediate processes

$$\tilde{\nu}_{\zeta,n}(y, \widehat{x}^*, x^*, z) = \sum_{i=1}^{n} \zeta_i \frac{(\mathbb{1}\{Y_i \leq y\} - F_{Y|X,Z}(y|X_i, z)) \cdot K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}}{\sqrt{nh_n} f_{X|Z}(x^*|z) \cdot Pr(Z = z)}.$$

It follows by Lemma 4 in Appendix B.1 that under Assumption 5,

$$\tilde{\nu}_{\zeta,n}(y, \widehat{x}^*, x^*, z) - \nu_{\zeta,n}(y, x^*, z) = o_P(1)$$

uniformly over (y, z). Finally, it follows by Lemma 5 in Appendix B.1 that

$$\widehat{\nu}_{\zeta,n}(y, \widehat{x}^*, z) - \tilde{\nu}_{\zeta,n}(y, \widehat{x}^*, x^*, z) = o_P(1)$$

uniformly over (y, z) under Assumption 5.

Combining the above results imply that $\widehat{\nu}_{\zeta,n}(y, \widehat{x}^*, z) \rightsquigarrow \mathbb{G}_z$. The result follows by the continuous mapping theorem as well as the property that $\|\mathbb{G}_z\|_\infty$ is continuously distributed (Kato, 2019, Theorem 22). $\square$

A.3 Theorem 3

First part For any $u \in \mathscr{C}$, there exists (x, z, z'), $z \neq z'$, such that $F_{X|Z}(x|z) = F_{X|Z}(x|z') = u$. Then,

$$\begin{aligned} E(Y|X = x, Z = z') - E(Y|X = x, Z = z) =& E(\varepsilon_0|\eta = u) + E(\varepsilon_1|\eta = u)x + E(\varepsilon_2|\eta = u)z' \\ & - (E(\varepsilon_0|\eta = u) + E(\varepsilon_1|\eta = u)x + E(\varepsilon_2|\eta = u)z) \\ =& E(\varepsilon_2|\eta = u)(z' - z). \end{aligned}$$

Therefore, $E(\varepsilon_2|\eta = u)$ is identified. Next, for any $u \in \mathscr{C}'$, let (x, z, z') be such that $F_{X|Z}(x|z') \neq F_{X|Z}(x|z) = u$. Then, because $F_{X|Z}(\cdot|z')$ is strictly increasing and continuous on $\text{Supp}(X|Z = z')$, there exists $x' \neq x$ such that $F_{X|Z}(x'|z') = u$. Moreover,

$$\begin{aligned}E(Y|X = x', Z = z') - E(Y|X = x, Z = z) =& E(\varepsilon_0|\eta = u) + E(\varepsilon_1|\eta = u)x' + E(\varepsilon_2|\eta = u)z' \\ & - (E(\varepsilon_0|\eta = u) + E(\varepsilon_1|\eta = u)x + E(\varepsilon_2|\eta = u)z) \\ =& E(\varepsilon_1|\eta = u)(x' - x) + E(\varepsilon_2|\eta = u)(z' - z).\end{aligned}$$

Hence, by what precedes, $E(\varepsilon_1|\eta = u)$ is identified.

Second part Let (x^*, z) be as in Assumption 4 and $u^* = F_{X|Z}(x^*|z)$. By the first part of the proof, we identify $E(\varepsilon_2|\eta = u^*)$. By Assumption 7, $E(\varepsilon_2) = E(\varepsilon_2|\eta = u) = E(\varepsilon_2|\eta = u^*)$ for all $u \in (0, 1)$, so we identify all these objects.

Next, let (z_1, z_2) be as in Assumption 8 and define $q_{z_1 z_2}(\cdot)$ as in Section 4.2 above. Because $F_{X|Z}(\cdot|z_2)$ is continuous, $F_{X|Z}(F^{-1}_{X|Z}(u|z_2)|z_2) = u$ for all $u \in (0, 1)$. Hence, $F_{X|Z}(q_{z_1 z_2}(x)|z_2) = F_{X|Z}(x|z_1)$ for all x such that $F_{X|Z}(x|z_1) \in (0, 1)$. Since for almost all x, $F_{X|Z}(x|z_1) \neq F_{X|Z}(x|z_2)$, this implies that for such x's, $q_{z_1 z_2}(x) \neq x$. Moreover, by Assumption 7 again,

$$\begin{aligned}E(Y - E(\varepsilon_2)Z|X = x, Z = z_1) =& E(\varepsilon_0|\eta = F_{X|Z}(x|z_1)) + E(\varepsilon_1|\eta = F_{X|Z}(x|z_1))x, \\ E(Y - E(\varepsilon_2)Z|X = q_{z_1 z_2}(x), Z = z_2) =& E(\varepsilon_0|\eta = F_{X|Z}(x|z_1)) + E(\varepsilon_1|\eta = F_{X|Z}(x|z_1))q_{z_1 z_2}(x).\end{aligned}$$

Therefore, we identify, for almost all $x \in \text{Supp}(X|Z = z_1)$, $E(\varepsilon_1|\eta = F_{X|Z}(x|z_1))$ by

$$E(\varepsilon_1|\eta = F_{X|Z}(x|z_1)) = \frac{E(Y - E(\varepsilon_2)Z|X = q_{z_1 z_2}(x), Z = z_2) - E(Y - E(\varepsilon_2)Z|X = x, Z = z_1)}{q_{z_1 z_2}(x) - x}.$$

This implies that $E(\varepsilon_1|\eta = u)$ is identified for almost all $u \in (0, 1)$, and then for all u by continuity of $u \mapsto E(\varepsilon_1|\eta = u)$ (in view of Assumption 1). Finally, $E(\varepsilon_1)$ is identified by $E(\varepsilon_1) = E(E(\varepsilon_1|\eta))$.

Third part Hereafter, (x^*, z, z') are defined in Assumption 4 and we assume without loss of generality that $|z'| > |z|$. We also let $u^* = F_{X|Z}(x^*|z)$. By Assumption 9, we have, for any $t \in \mathbb{R}$,

$$\Psi_{Y|\eta=u^*, Z=z'}(t) = \Psi_{\varepsilon_0+\varepsilon_1 x^*|\eta=u^*}(t)\Psi_{\varepsilon_2}(z't), \tag{A.3}$$

which is different from 0 by Assumption 10 (ii). Then

$$\frac{\Psi_{Y|X=x^*,Z=z'}(t/z')}{\Psi_{Y|X=x^*,Z=z}(t/z')} = \frac{\Psi_{\varepsilon_2}(t)}{\Psi_{\varepsilon_2}(st)},$$

with $s = z/z'$. In other words, we identify the function $R(t) = \Psi_{\varepsilon_2}(t)/\Psi_{\varepsilon_2}(st)$. Hence, for any integer K, we also identify $\prod_{k=0}^{K} R(s^k t) = \Psi_{\varepsilon_2}(t)/\Psi_{\varepsilon_2}(s^{K+1}t)$, with $|s| < 1$. By the continuity of Ψ_{ε_2} at 0, we in turn identify $\Psi_{\varepsilon_2}(t)$ by $\prod_{k=0}^{\infty} R(s^k t)$.

Next, for all $(t, u, z) \in \mathbb{R}^* \times (0, 1) \times \mathcal{Z}$, using again the fact that Ψ_{ε_2} does not vanish on the real line, we have

$$\Psi_{\varepsilon_0,\varepsilon_1|\eta}(t, th(z, u)|u) = \frac{\Psi_{Y|\eta=u,Z=z}(t)}{\Psi_{\varepsilon_2}(zt)}.$$

Therefore, $\Psi_{\varepsilon_0,\varepsilon_1|\eta}(t, th(z, u)|u)$ is identified for all $(t, u, z) \in \mathbb{R}^* \times (0, 1) \times \mathcal{Z}$. In other words, for all $(t, u) \in \mathbb{R}^* \times (0, 1)$, the function $S_{t,u} : t' \mapsto \Psi_{\varepsilon_0,\varepsilon_1|\eta}(t, t'|u)$ is identified on $\text{Supp}(th(Z, u))$. By Assumption 10 (i), this set admits a limit point. Moreover, by the dominated convergence theorem and Assumption 10 (iii), $S_{t,u}$ is analytic on the strip $\mathcal{A} = \{z \in \mathbb{C} : |\text{Im}(z)| < c_1\}$ for all $(t, u) \in \mathbb{R}\times(0, 1)$. This implies that the function $S_{t,u}$ admits a unique analytic continuation from $\text{Supp}(th(Z, u))$ to $\mathcal{A}$. As a result, $S_{t,u}$ is identified on $\mathbb{R}$. Hence, we identify $\Psi_{\varepsilon_0,\varepsilon_1|\eta}(t, t'|u)$ for all $(t, t', u) \in \mathbb{R}^* \times \mathbb{R} \times (0, 1)$. Finally, by the continuity of $t \mapsto \Psi_{\varepsilon_0,\varepsilon_1|\eta}(t, t'|u)$, $\Psi_{\varepsilon_0,\varepsilon_1|\eta}(0, \cdot|\cdot)$ is also identified on $\mathbb{R} \times (0, 1)$. The result follows. □

A.4 Theorem 4

Suppose without loss of generality that $z' > z$. Hereafter, we let $\text{sgn}(x) = x/|x|$ if $x \neq 0$, 0 otherwise. By Assumption 11 (ii), the map $w \mapsto E(g(w, \varepsilon)|X = x, Z = z)$ is strictly increasing. Therefore, using Assumption 11 (i),

$$\text{sgn}\left(E\left[g(x^*\alpha_0 + z'\beta_0, \varepsilon) - g(x^*\alpha_0 + z\beta_0, \varepsilon)|X = x^*, Z = z'\right]\right) = \text{sgn}(\beta_0) = \beta_0. \quad \text{(A.4)}$$

Moreover, we have that

$$
\begin{aligned}
E(Y|X=x^*,Z=z) &= E\left[g(x^*\alpha_0+z\beta_0,\varepsilon)|X=x^*,Z=z\right] \\
&= E\left[g(x^*\alpha_0+z\beta_0,\varepsilon)|\eta=F_{X|Z}(x^*|z),Z=z\right] \\
&= E\left[g(x^*\alpha_0+z\beta_0,\varepsilon)|\eta=F_{X|Z}(x^*|z),Z=z'\right] \\
&= E\left[g(x^*\alpha_0,+z\beta_0\varepsilon)|X=q_{zz'}(x^*),Z=z'\right] \\
&= E\left[g(x^*\alpha_0+z\beta_0,\varepsilon)|X=x^*,Z=z'\right]. \qquad \text{(A.5)}
\end{aligned}
$$

The first equality stems from Assumption 11. The second is due to monotonicity of the first stage (Assumption 3). The third follows from exogeneity of the instrument (Assumption 2). The fourth is due to monotonicity of the first stage again. The last follows by noting that, by Assumption 4, $q_{zz'}(x^*)=x^*$.

By combining (A.4) and (A.5), we obtain that β_0 is identified through

$$
\beta_0=\operatorname{sgn}\left[E\left(Y|X=x^*,Z=z'\right)-E\left(Y|X=x^*,Z=z\right)\right]. \qquad \text{(A.6)}
$$

Next, let us turn to α_0. Suppose first that $E\left[Y|X=x^*,Z=z\right]=E\left[Y|X=x^*,Z=z'\right]$, or equivalently, $\beta_0=0$. In this case, $\alpha_0\in\{-1,0,1\}$. We can use the same reasoning as above to show that

$$
\alpha_0=\frac{\operatorname{sgn}\left[E\left(Y|X=q_{z_1z_1'}(x_1),Z=z_1'\right)-E\left(Y|X=x_1,Z=z_1\right)\right]}{\operatorname{sgn}\left[q_{z_1z_1'}(x_1)-x_1\right]}.
$$

Hence, α_0 is identified as well in this case.

Consider the second case where $E\left[Y|X=x^*,Z=z\right]\neq E\left[Y|X=x^*,Z=z'\right]$, or, equivalently, $\beta_0\neq 0$. We have, with the same reasoning as that used to obtain (A.5),

$$
\begin{aligned}
E(Y|X=x_1,Z=z_1) &= E\left[g(x_1\alpha_0+z_1'\beta_0,\varepsilon)|X=x_1,Z=z_1\right] \\
&= E\left[g(x_1\alpha_0+z_1\beta_0,\varepsilon)|\eta=h^{-1}(z_1',x_1),Z=z_1'\right] \\
&= E\left[g(x_1\alpha_0+z_1\beta_0,\varepsilon)|X=q_{z_1z_1'}(x_1),Z=z_1'\right].
\end{aligned}
$$

The same holds for $E(Y|X = q_{z_1z_1'}(x_1), Z = z_1')$. Because $E(Y|X = x_1, Z = z_1) = E(Y|X = q_{z_1z_1'}(x_1), Z = z_1')$, by Assumption 12 (ii), we have

$$E\left[g(x_1\alpha_0 + z_1\beta_0, \varepsilon)|X = q_{z_1z_1'}(x_1), Z = z_1'\right] = E\left[g(q_{z_1z_1'}(x_1)\alpha_0 + z_1'\beta_0, \varepsilon)|X = q_{z_1z_1'}(x_1), Z = z_1'\right].$$

Finally, by Assumption 11 (ii),

$$x_1\alpha_0 + z_1\beta_0 = q_{z_1z_1'}(x_1)\alpha_0 + z_1'\beta_0.$$

Because $q_{z_1z_1'}(x_1) \neq x_1$, α_0 is identified by $\alpha_0 = (z_1 - z_1')\beta_0/(q_{z_1z_1'}(x_1) - x_1)$. □

A.5 Theorem 5

Let $e(w, u) = E(Y|W = w, \eta = u)$. Because W is identified under Assumptions 1-4 and 11-12, by Theorem 4, $e(\cdot, \cdot)$ is identified on $\text{Supp}(W, \eta)$. Moreover, under Assumption 13, we can identify, for almost all $u \in (0, 1)$ and all $w \in \text{Supp}(W|\eta = u)$, $e(w_n, u)$ on a sequence $(w_n)_{n\in\mathbb{N}}$ tending to w. As a result, we can identify $\partial_w e(w, u)$. Now, note that

$$\Delta_X(X, Z) = \alpha_0 \partial_w e(W, \eta),$$

$$\Delta_Z(X, Z) = \beta_0 \partial_w e(W, \eta).$$

As a result, we identify $\Delta_X(X, Z)$ and $\Delta_Z(X, Z)$. Moreover, $\Delta_X = \alpha_0 E\left[\partial_w e(W, \eta)\right]$ and $\Delta_Z = \beta_0 E\left[\partial_w e(W, \eta)\right]$. Hence, Δ_X and Δ_Z are also identified. □

A.6 Theorem 6

First part. Hereafter, we denote by $\text{Int}(A)$ the interior of any set A in a topological space. We first show that, for all $z \in \mathcal{Z}$, $g_1(z, \cdot)$ is identified on $\text{Supp}(Y|Z = z_1)$, where z_1 is defined in Assumption 15 (ii). For that purpose, we first prove that $g_1(z_0, \cdot)$ is identified on $\text{Int}(\text{Supp}(Y|Z = z_1))$. By Assumption 16, z_1 is connected with z_0. Therefore, there exist $n \in \mathbb{N}$ and $(x^k, z^k)_{k=0,\ldots,n}$, with $z^0 = z_0$ and $z^n = z_1$, such that $F_{X|Z}(x^k|z^k) = F_{X|Z}(x^k|z^{k+1})$ for all $k = 0, ..., n-1$. Let us show by the reverse induction on k that, for all $k = 0, ..., n$ and all $u \in \text{Int}(\text{Supp}(Y|Z = z_1)$, $g_1(z^k, u)$ is identified and $g_1(z^k, u) \in$

Int(Supp($Y|Z = z^k$)). By Assumption 15 (ii), $g_1(z^n, u) = u$ for all $u \in$ Int(Supp($Y|Z = z_1$)), so the result holds for $k = n$. Let us show that, if it holds for $k + 1 \in \{1, ..., n\}$, then it also holds for k. By Equation (A.2) and strict monotonicity of F_η on Supp(η), we have $h^{-1}(z^k, x^k) = h^{-1}(z^{k+1}, x^k)$. Moreover, for all $u \in \mathbb{R}$,

$$\begin{aligned} F_{Y|X,Z}\left(g_1(z^k, u)|x^k, z^k\right) &= \Pr\left(g_1(z^k, \varphi(x^k, \varepsilon)) \le g_1(z^k, u)|Z = z^k, \eta = h^{-1}(z^k, x^k)\right) \\ &= \Pr\left(\varphi(x^k, \varepsilon) \le u|\eta = h^{-1}(z^{k+1}, x^k)\right) \\ &= \Pr\left(g_1(z^{k+1}, \varphi(x^k, \varepsilon)) \le g_1(z^{k+1}, u)|Z = z^{k+1}, \eta = h^{-1}(z^{k+1}, x^k)\right) \\ &= F_{Y|X,Z}\left(g_1(z^{k+1}, u)|x^k, z^{k+1}\right). \end{aligned} \tag{A.7}$$

The first equality follows by Assumptions 3 and 15. The second is due to Assumptions 2 and 15 again. The third follows by the same reasoning, and using $h^{-1}(z^k, x_k) = h^{-1}(z^{k+1}, x_k)$. The fourth equality is obtained as the first one. Now, by the induction hypothesis and because Supp($Y|X = x^k, Z = z^{k+1}$) = Supp($Y|Z = z^{k+1}$), we have, for all $u \in$ int(Supp($Y|Z = z_1$)),

$$F_{Y|X,Z}\left(g_1(z^k, u)|x^k, z^k\right) = F_{Y|X,Z}\left(g_1(z^{k+1}, u)|x^k, z^{k+1}\right) \in (0, 1).$$

By Assumption 14, Supp($Y|Z = z^k$) = Supp($Y|X = x^k, Z = z^k$) is an interval. Thus, $g_1(z^k, u) \in$ Int(Supp($Y|Z = z^k$)), on which $F_{Y|X,Z}$ is strictly increasing. Therefore, by (A.7),

$$g_1(z^k, u) = F^{-1}_{Y|X,Z}\left[F_{Y|X,Z}\left(g_1(z^{k+1}, u)|x^k, z^{k+1}\right)|x^k, z^k\right].$$

Hence, $g_1(z^k, u)$ is identified, and the property holds for k. As a result, it holds for all $k = 0, ..., n$. Thus, $g_1(z_0, \cdot)$ is identified on Int(Supp($Y|Z = z_1$)). Next, consider $z \in \mathcal{Z}$. z is connected with z_0 by Assumption 16. The same reasoning as above implies that $g_1(z, \cdot)$ is identified on Int(Supp($Y|Z = z_1$)). Finally, by continuity of $g_1(z, \cdot)$ on the boundary of Supp($Y|Z = z_1$), $g_1(z, \cdot)$ is identified on Supp($Y|Z = z_1$).

We now prove that average effects are identified. For that purpose, let $g_1^{-1}(z, \cdot)$ denote the inverse of $g_1(z, \cdot)$ and let us define $Y_z = g_1(z, g_1^{-1}(Z, Y))$ for any $z \in \mathcal{Z}$. We have $g_1^{-1}(Z, Y) = \varphi(X, \varepsilon)$ and by

Assumption 15 (ii)–(iii),

$$\text{Supp}(\varphi(X,\varepsilon)|Z=z)=\text{Supp}(\varphi(X,\varepsilon)|Z=z_1)=\text{Supp}(Y|Z=z_1).$$

Thus, by what precedes, Y_z is identified for all $z \in \mathcal{Z}$. Then, for all (x, z, z') such that $(x, z) \in \text{Supp}(X, Z)$ and $z' \in \mathcal{Z}$,

$$E(Y_{z'}-Y|X=x,Z=z)=E(g_1(z',\varphi(x,\varepsilon))-g_1(z,\varphi(x,\varepsilon))|X=x,Z=z)=\Delta_Z(x,z,z'),$$

which shows that $\Delta_Z(x, z, z')$ is identified. Next,

$$\begin{aligned}
&E(Y_z|X=q_{zz'}(x),Z=z')-E(Y|X=x,Z=z)\\
=&E\left[g_1(z,\varphi(q_{zz'}(x),\varepsilon))|\eta=h^{-1}(z',q_{zz'}(x))\right]-E(Y|X=x,Z=z)\\
=&E\left[g_1(z,\varphi(q_{zz'}(x),\varepsilon))|\eta=h^{-1}(z,x)\right]-E(Y|X=x,Z=z)\\
=&E\left[g_1(z,\varphi(q_{zz'}(x),\varepsilon))-g_1(z,\varphi(x,\varepsilon))|X=x,Z=z\right]\\
=&\Delta_X(x,q_{zz'}(x),z),
\end{aligned}$$

which implies that $\Delta_X(x, q_{zz'}(x), z)$ is identified.

Second part. As indicated in the statement of the theorem, we also rely on the following regularity conditions. Condition (iii) is imposed to ensure that expectations and derivative can be interchanged. It could be weakened but at the price of complicating the condition.

Assumption 18 (Regularity conditions)**.** *(i) $h(\cdot, u)$ is continuous for almost all $u \in (0, 1)$.*
(ii) g_1 and $\varphi(\cdot, e)$ are differentiable; for almost all $e \in Supp(\varepsilon)$.
(iii) there exist positive functions $s(\cdot)$, $t(\cdot)$ and $u(\cdot)$ such that for almost every (z, u, e), $|\partial_z g_1(z, u)| \leq s(u)$, $|\partial_u g_1(z, u)| \leq t(z)$ and $|\partial_x \varphi(x, e)| \leq u(e)$ with $E[s(\varphi(X, \varepsilon))] < +\infty$ and $E[t(Z)] < +\infty$ and $E[u(\varepsilon)] < +\infty$.

Now, by Assumption 17, there exists, for all $(x, z, n) \in \text{Supp}(X, Z) \times \mathbb{N}$, an identified function $b_n(\cdot)$ such that $b_n(z) \in \mathcal{Z}$ and $0 < |b_n(z) - z| < 1/(n + 1)$. Then

$$E\left[\frac{Y_{b_n(z)}-Y}{b_n(z)-z}\middle|X=x,Z=z\right]=E\left[\frac{g_1(b_n(z),\varphi(x,\varepsilon))-g_1(z,\varphi(x,\varepsilon))}{b_n(z)-z}\middle|X=x,Z=z\right].$$

Moreover, by Assumption 18 (iii),

$$\left|\frac{g_1(b_n(z), \varphi(x,\varepsilon)) - g_1(z, \varphi(x,\varepsilon))}{b_n(z) - z}\right| \leq s(\varphi(x,\varepsilon))$$

with $E(s(\varphi(x,\varepsilon))|X = x, Z = z) < +\infty$ for almost all (x, z). Thus, by the dominated convergence theorem, as $n \to \infty$,

$$E\left[\frac{Y_{b_n(z)} - Y}{b_n(z) - z}\middle| X = x, Z = z\right] \to E\left[\partial_z g_1(z, \varphi(x,\varepsilon))\middle| X = x, Z = z\right] = \Delta_Z(x, z).$$

Because this holds for almost all $(x, z) \in \text{Supp}(X, Z)$, $\Delta_Z = E[\Delta_Z(X, Z)]$ is also identified.

We now show that $\Delta_X(x, z)$ is identified for almost every (x, z). By Assumption 18 (i), for all $n \in \mathbb{N}$ and almost all $(x, z) \in \text{Supp}(X, Z)$ with $z \in \text{Int}(\mathcal{Z})$, there exists $\delta_n > 0$ such that any $z' \in \mathcal{Z}$ satisfying $|z' - z| < \delta_n$ will be such that $|h(z', h^{-1}(z, x)) - h(z, h^{-1}(z, x))| < 1/(n + 1)$. Moreover, by Assumption 17 applied to the neighborhood $\mathcal{V} = (z - \delta_n, z + \delta_n) \cap \text{Int}(\mathcal{Z})$, there exists an identified function $r_n(\cdot)$ such that $r_n(z) \in \mathcal{Z}$ and $h(r_n(z), h^{-1}(z, x)) \neq h(z, h^{-1}(z, x)) = x$. By definition, $h(z', h^{-1}(z, x)) = q_{zz'}(x)$. Therefore,

$$0 < |q_{zr_n(z)}(x) - z| < \frac{1}{n+1}.$$

Now, let us define, for all $n \in \mathbb{N}$ and $(x, z) \in \text{Supp}(X, Z)$ with $z \in \text{Int}(\mathcal{Z})$,

$$\lambda_n(x, z) = \frac{E[Y_z|X = q_{zr_n(z)}(x), Z = r_n(z)] - E[Y_z|X = x, Z = z]}{q_{zr_n(z)}(x) - x}.$$

Reasoning as above, we obtain

$$\lambda_n(x, z) = E\left[\frac{g_1(z, \varphi(q_{zr_n(z)}(x), \varepsilon)) - g_1(z, \varphi(x, \varepsilon))}{q_{zr_n(z)}(x) - x} | X = x, Z = z\right].$$

Moreover, by Assumption 18 (iii),

$$\left|\frac{g(q_{zr_n(z)}(x), z, \varepsilon)) - g(x, z, \varepsilon)}{q_{zr_n(z)}(x) - x}\right| \leq t(z)r(\varepsilon),$$

with $E[r(\varepsilon)|X = x, Z = z] < +\infty$. Hence, by the dominated convergence theorem, as $n \to \infty$,

$$\Delta_X(x, z) = \lim_{n\to\infty} \lambda_n(x, z).$$

and $\Delta_X(x, z)$ is identified. Since this holds for almost all $(x, z) \in \text{Supp}(X, Z)$, Δ_X is also identified. $\square$

B Auxiliary Lemmas

B.1 Auxiliary Lemmas used in the proof of Theorem 2

This section lists a number of auxiliary lemmas to prove Theorem 2.

Lemma 1. *Suppose that Assumption 5 (i)–(ii) holds. Then,* $\widehat{x}^* - x^* = o_P(1)$.

Lemma 2. *Suppose that Assumption 5 (i)–(ii) holds. Then,* $\widehat{x}^* - x^* = O_P(n^{-1/2})$.

Lemma 3. *Suppose that Assumption 5 holds. Then,*

$$\sqrt{nh_n}(\widehat{F}_{Y|X,Z}(y|x^*,z) - F_{Y|X,Z}(y|x^*,z)) = \nu_n(y,x^*,z) + o_P(1)$$

uniformly over (y,z).

Lemma 4. *Suppose that Assumption 5 holds. Then*

$$\sqrt{nh_n}(\widehat{F}_{Y|X,Z}(y|\widehat{x}^*,z) - \widehat{F}_{Y|X,Z}(y|x^*,z)) = o_P(1) \tag{B.1}$$

uniformly over (y,z), *and*

$$\tilde{\nu}_{\zeta,n}(y,\widehat{x}^*,x^*,z) - \nu_{\zeta,n}(y,x^*,z) = o_P(1) \tag{B.2}$$

uniformly over (y,z), *where*

$$\tilde{\nu}_{\zeta,n}(y,\widehat{x}^*,x^*,z) = \sum_{i=1}^{n} \zeta_i \frac{(\mathbb{1}\{Y_i \leq y\} - F_{Y|X,Z}(y|X_i,z)) \cdot K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}}{\sqrt{nh_n} f_{X|Z}(x^*|z) \cdot Pr(Z=z)}.$$

Lemma 5. *Suppose that Assumption 5 holds. Then,*

$$\widehat{\nu}_{\zeta,n}(y,\widehat{x}^*,z) - \tilde{\nu}_{\zeta,n}(y,\widehat{x}^*,x^*,z) = o_P(1)$$

uniformly over (y,z).

B.2 Proof of Lemma 1

Proof. Let $M_n(x) = -|\widehat{F}_{X|Z}(\cdot|0) - \widehat{F}_{X|Z}(\cdot|1)|$, $M(x) = -|F_{X|Z}(\cdot|0) - F_{X|Z}(\cdot|1)|$ and

$$\widehat{I} = [\widehat{F}^{-1}_{X|Z}(\underline{p}|0), \widehat{F}^{-1}_{X|Z}(\overline{p}|0)].$$

By Assumption 5 (i)–(ii), there exists $\xi > 0$ such that $x^* \in I_\xi = (F^{-1}_{X|Z}(\underline{p}|0) + \xi, F^{-1}_{X|Z}(\overline{p}|0) - \xi)$. Fix $\kappa > 0$ and let $B = \{x \in I_\xi : \ |x - x^*| > \kappa\}$. Since the data are i.i.d., by Glivenko-Cantelli Theorem,

$$\begin{aligned} \|M_n - M\|_\infty &\leq \|\widehat{F}_{X|Z=0} - F_{X|Z=0}\|_\infty + \|\widehat{F}_{X|Z=1} - F_{X|Z=1}\|_\infty \\ &= o_P(1). \end{aligned} \tag{B.3}$$

Now, B is included in a compact set $\widetilde{B}$ that does not include x^*. Moreover, M is continuous on $\widetilde{B}$ by Assumption 5 (ii). Then, by Assumption 5 (i), $\sup_{x\in B} M(x) \leq \max_{x\in\widetilde{B}} M(x) < M(x^*) = 0$. Thus,

$$\begin{aligned} \sup_{x\in B} M_n(x) \leq & \|M_n - M\|_\infty + \sup_{x\in B} M(x) \\ < & \|M_n - M\|_\infty. \end{aligned} \tag{B.4}$$

Note that if $\widehat{x}^* \in B$ and $I_\xi \subset \widehat{I}$, then $M_n(x^*) \leq M_n(\widehat{x}^*) = \sup_{x\in B} M_n(x)$. Therefore,

$$\begin{aligned} \Pr\left(\widehat{x}^* \in B \text{ and } I_\xi \subset \widehat{I}\right) \leq & \Pr\left(\sup_{x\in B} M_n(x) - M_n(x^*) \geq 0\right) \\ \to & 0, \end{aligned}$$

where the last line follows from (B.3)–(B.4). By the law of large numbers, $\Pr(I_\xi \subset \widehat{I}) \to 1$. Hence, $\Pr(\widehat{x}^* \in B) \to 0$. For κ small enough, $\widehat{x}^* \notin I_\xi$ implies that $|\widehat{x}^* - x^*| > \kappa$ and then $\Pr(|\widehat{x}^* - x^*| > \kappa) \to 0$. The result follows. ☐

B.3 Proof of Lemma 2

Proof. For short-hand notations, let $p^* = (p_0^*, p_1^*) = (\Pr(Z = 0), \Pr(Z = 1))$ and $\widehat{p}^* = (\widehat{p}_0^*, \widehat{p}_1^*) = (\widehat{\Pr}(Z = 0), \widehat{\Pr}(Z = 1))$, where $\widehat{\Pr}(Z = 0) = \mathbb{E}_n[\mathbb{1}\{Z = 0\}]$ and $\widehat{\Pr}(Z = 1) = \mathbb{E}_n[\mathbb{1}\{Z = 1\}]$. For a generic notation, let $p \in (0,1)^2$. Also let

$$\widehat{\Psi}(x,p) = p_0^{-1} \cdot \mathbb{E}_n[\mathbb{1}\{u \leq x\} \cdot \mathbb{1}\{v = 0\}] - p_1^{-1} \cdot \mathbb{E}_n[\mathbb{1}\{u \leq x\} \cdot \mathbb{1}\{v = 1\}],$$

$$\varphi_{x,p}(u,v) = p_0^{-1} \cdot \mathbb{1}\{u \leq x\} \cdot \mathbb{1}\{v = 0\} - p_1^{-1} \cdot \mathbb{1}\{u \leq x\} \cdot \mathbb{1}\{v = 1\},$$

and $\Psi(x,p) = E[\varphi_{x,p}(X,Z)]$. Then, $\mathcal{F}_\delta = \{\varphi_{x,p} : \ |x - x^*| \vee |p_0 - p_0^*| \vee |p_1 - p_1^*| < \delta\}$ is Donsker for any $\delta > 0$, and it holds that $\Psi(x^*, p^*) = 0$. Since $0 \neq 1$, we have

$$\begin{aligned} E\left[(\varphi_{x,p^*}(X,Z) - \varphi_{x^*,p^*}(X,Z))^2\right] =& (p_0^*)^{-1} F_{X|Z}(x|0) + (p_1^*)^{-1} F_{X|Z}(x|1) \\ &+ (p_0^*)^{-1} F_{X|Z}(x^*|0) + (p_1^*)^{-1} F_{X|Z}(x^*|1) \\ &- 2(p_0^*)^{-1} F_{X|Z}(x \wedge x^*|0) - 2(p_1^*)^{-1} F_{X|Z}(x \wedge x^*|1). \end{aligned}$$

Therefore, by Assumption 5 (ii),

$$E\left[(\varphi_{x,p^*}(X,Z) - \varphi_{x^*,p^*}(X,Z))^2\right] \to 0 \qquad \text{as } x \to x^*.$$

This result, together with Lemma 1 and Lemma 19.24 of van der Vaart (1998), implies that

$$\sqrt{n}\left(\widehat{\Psi}(\widehat{x}^*, p^*) - \Psi(\widehat{x}^*, p^*) - \widehat{\Psi}(x^*, p^*) + \Psi(x^*, p^*)\right) = o_P(1).$$

Note that $\Psi(x^*, p^*) = 0$ holds. Therefore, the central limit theorem yields $\widehat{\Psi}(x^*, p^*) = O_P(n^{-1/2})$, $\widehat{p}_0^* - p_0^* = O_P(n^{-1/2})$ and $\widehat{p}_1^* - p_1^* = O_P(n^{-1/2})$, and hence $\widehat{\Psi}(x^*, \widehat{p}^*) = O_P(n^{-1/2})$. As $|\widehat{\Psi}(\widehat{x}^*, \widehat{p}^*)| \leq |\widehat{\Psi}(x^*, \widehat{p}^*)|$, these rates also imply $\widehat{\Psi}(\widehat{x}^*, \widehat{p}^*) = O_P(n^{-1/2})$. Furthermore, $\widehat{p}_0^* - p_0^* = O_P(n^{-1/2})$ and $\widehat{p}_1^* - p_1^* = O_P(n^{-1/2})$ imply that $\|\widehat{\Psi}(\,\cdot\,, \widehat{p}^*) - \widehat{\Psi}(\,\cdot\,, p^*)\|_\infty = O_P(n^{-1/2})$, which in turn implies

$$\begin{aligned} \sqrt{n}\widehat{\Psi}(\widehat{x}^*, p^*) &= \sqrt{n}\widehat{\Psi}(\widehat{x}^*, \widehat{p}^*) + \sqrt{n}(\widehat{\Psi}(\widehat{x}^*, p^*) - \widehat{\Psi}(\widehat{x}^*, \widehat{p}^*)) \\ &= O_P(1). \end{aligned}$$

Using all these auxiliary rate results together, we have

$$\begin{aligned} \sqrt{n}\left(\Psi(\widehat{x}^*, p^*) - \Psi(x^*, p^*)\right) =& \sqrt{n}\widehat{\Psi}(\widehat{x}^*, p^*) - \sqrt{n}\widehat{\Psi}(x^*, p^*) \\ &- \sqrt{n}\left(\widehat{\Psi}(\widehat{x}^*, p^*) - \Psi(\widehat{x}^*, p^*) - \widehat{\Psi}(x^*, p^*) + \Psi(x^*, p^*)\right) \\ =& O_P(1). \end{aligned} \tag{B.5}$$

Finally, by Assumption 5 (ii), $\Psi(\cdot, p^*) = F_{X|Z}(\cdot|1) - F_{X|Z}(\cdot|0)$ is differentiable. Thus, by the mean value theorem, there exists $\widetilde{x}^* \in (x^* \wedge \widehat{x}^*, x^* \vee \widehat{x}^*)$ such that

$$\Psi(\widehat{x}^*, p^*) - \Psi(x^*, p^*) = \left(f_{X|Z}(\widetilde{x}^*|1) - f_{X|Z}(\widetilde{x}^*|0)\right)(\widehat{x}^* - x^*).$$

By Assumption 5 (ii), there exists a neighborhood $\mathcal{V}$ of x^* and $C > 0$ such that

$$\inf_{x \in \mathcal{V}} \left|f_{X|Z}(x|1) - f_{X|Z}(x|0)\right| > C.$$

Moreover, by Lemma 1, $\widehat{x}^* \in \mathcal{V}$ with probability tending to one. Then, under this event,

$$\begin{aligned}
|\Psi(\widehat{x}^*, p^*) - \Psi(x^*, p^*)| =& \left|f_{X|Z}(\widetilde{x}^*|1) - f_{X|Z}(\widetilde{x}^*|0)\right| |\widehat{x}^* - x^*| \\
\geq& C|\widehat{x}^* - x^*|.
\end{aligned}$$

Combined with (B.5), this implies that $\widehat{x}^* - x^* = O_P(n^{-1/2})$. □

B.4 Proof of Lemma 3

Proof. The mean value expansion under Assumption 5 (iii) yields

$$\begin{aligned}
\mathbb{1}\{Y_i \leq y\} =& F_{Y|X,Z}(y|X_i, z) + \mathbb{1}\{Y_i \leq y\} - F_{Y|X,Z}(y|X_i, z) \\
=& F_{Y|X,Z}(y|x^*, z) + \partial_x F_{Y|X,Z}(y|\iota(X_i, x^*), z) \cdot (X_i - x^*) + \mathbb{1}\{Y_i \leq y\} - F_{Y|X,Z}(y|X_i, z),
\end{aligned}$$

where $\iota(X_i, x^*)$ is some point between X_i and x^*. Substituting this equation in the closed-form expression of $\widehat{F}_{Y|X,Z}(y|x^*, z)$ yields

$$\sqrt{nh_n}\widehat{F}_{Y|X,Z}(y|x^*, z) = \sqrt{nh_n}F_{Y|X,Z}(y|x^*, z) + \frac{A_1(y, z) + A_2(y, z)}{A_3(z)}, \tag{B.6}$$

where

$$\begin{aligned}
A_1(y, z) =& \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \partial_x F_{Y|X,Z}(y|\iota(X_i, x^*), z) \cdot (X_i - x^*) \cdot K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}, \\
A_2(y, z) =& \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n [\mathbb{1}\{Y_i \leq y\} - F_{Y|X,Z}(y|X_i, z)] \cdot K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}, \qquad \text{and} \\
A_3(z) =& \frac{1}{nh_n} \sum_{i=1}^n K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}.
\end{aligned}$$

For convenience of presentation, we also introduce the following notation.

$$A_1^*(y,z) = \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \partial_x F_{Y|X,Z}(y|x^*,z) \cdot (X_i - x^*) \cdot K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}.$$

First, under Assumption 5, we obtain the deterministic order

$$\begin{aligned} E[A_1^*(y,z)] &= \sqrt{nh_n^3} \Pr(Z=z) \cdot \partial_x F_{Y|X,Z}(y|x^*,z) \cdot \int t \cdot K(t) \cdot f_{X|Z}(x^* + th_n|z)dt \\ &= O\left(\sqrt{nh_n^3}\right) \end{aligned}$$

uniformly over (y,z). Letting $\Delta_i = \partial_x F_{Y|X,Z}(y|\iota(X_i^*,x^*),z) - \partial_x F_{Y|X,Z}(y|x^*,z)$, we also have

$$\begin{aligned} E[(A_1(y,z) - A_1^*(y,z))^2] &= \frac{n}{nh_n} E\left[\Delta_i^2 (X_i - x^*)^2 K_h^2(X_i - x^*) \mathbb{1}\{Z_i = z\}\right] \\ &\quad + \frac{n(n-1)}{nh_n} E\left[\Delta_i (X_i - x^*) K_h(X_i - x^*) \mathbb{1}\{Z_i = z\}\right]^2 \\ &\lesssim \frac{1}{h_n} E\left[(X_i - x^*)^4 K_h^2(X_i - x^*)\right] + \frac{n}{h_n} E\left[(X_i - x^*)^2 |K_h(X_i - x^*)|\right]^2 \\ &\lesssim h_n^4 + nh_n^5 \lesssim nh_n^5 \end{aligned}$$

uniformly over (y,z) under Assumption 5 (iii)–(v). Also,

$$\begin{aligned} V(A_1^*(y,z)) &\leq h_n^2 \partial_x F_{Y|X,Z}(y|x^*,z) \cdot \int t^2 \cdot K(t) \cdot f_{X|Z}(x^* + th_n|z)dt \\ &= O(h_n^2) \end{aligned}$$

uniformly over (y,z) under Assumption 5 (ii)–(v). It thus follows that

$$\begin{aligned} E[A_1(y,z)^2] &= E\left[\left((A_1(y,z) - A_1^*(y,z)) + (A_1^*(y,z) - E[A_1^*(y,z)]) + E[A_1^*(y,z)]\right)^2\right] \\ &\lesssim E\left[(A_1(y,z) - A_1^*(y,z))^2\right] + V(A_1^*(y,z)) + E[A_1^*(y,z)]^2 \\ &\lesssim nh_n^5 + h_n^2 + nh_n^3 = o(1) \end{aligned}$$

uniformly over (y,z) under Assumption 5, and this implies $A_1(y,z) = o_P(1)$ uniformly over (y,z).

Second, similarly to the calculations above, we have

$$\begin{aligned} V(A_2(y,z)) &\leq \int K(t) \cdot f_{X|Z}(x^* + th_n|z)dt \\ &= O(1) \end{aligned}$$

uniformly over (y,z) under Assumption 5 (ii), (iv) and (v), and this implies $A_2(y,z) = O_P(1)$ uniformly over (y,z).

Third, $E[A_3(z)] - f_{X|Z}(x^*|z) \cdot \Pr(Z=z) = O(h_n)$ uniformly over z under Assumption 5 (ii), (iv), and (v). Also, $A_3(z) - E[A_3(z)] = O_P\left(\frac{1}{\sqrt{nh_n}}\right)$ uniformly over z under Assumption 5 (iv)–(v). It thus follows that

$$
\begin{aligned}
A_3(z) - f_{X|Z}(x^*|z) \cdot \Pr(Z=z) &= O(h_n) + O_P\left(\frac{1}{\sqrt{nh_n}}\right) \\
&= o_P(1)
\end{aligned}
$$

uniformly over z under Assumption 5 (iv). By Assumption 5 (ii), there exists a neighborhood $\mathcal{V}$ of x^* and $C > 0$ such that $\inf_{x \in \mathcal{V}} \left| f_{X|Z}(x|1) - f_{X|Z}(x|0) \right| > C$. By the continuous mapping theorem, $1/A_3(z) - 1/[\Pr(Z=z) \cdot f_{X|Z}(x^*|z)] = o_P(1)$ uniformly over z under Assumption 5 (ii).

Finally, the claim of the lemma follows by (B.6). □

B.5 Proof of Lemma 4

Proof. Note that we can write

$$
\begin{aligned}
&\sqrt{nh_n}(\widehat{F}_{Y|X,Z}(y|\widehat{x}^*,z) - \widehat{F}_{Y|X,Z}(y|x^*,z)) \\
=&\sqrt{nh_n}\frac{A_3(z)\left[\widehat{A}_4(y,z) - A_4(y,z)\right] - A_4(y,z)\left[\widehat{A}_3(z) - A_3(z)\right]}{A_3(z)\left[\widehat{A}_3(z) - A_3(z)\right] + A_3(z)^2},
\end{aligned} \tag{B.7}
$$

where

$$
\begin{aligned}
A_3(z) =& \frac{1}{nh_n}\sum_{i=1}^{n} K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\}, \\
A_4(y,z) =& \frac{1}{nh_n}\sum_{i=1}^{n} \mathbb{1}\{Y_i \le y\} \cdot K_{h_n}(X_i - x^*) \cdot \mathbb{1}\{Z_i = z\},
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{A}_3(z) =& \frac{1}{nh_n}\sum_{i=1}^{n} K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}, \\
\widehat{A}_4(y,z) =& \frac{1}{nh_n}\sum_{i=1}^{n} \mathbb{1}\{Y_i \le y\} \cdot K_{h_n}(X_i - \widehat{x}^*) \cdot \mathbb{1}\{Z_i = z\}.
\end{aligned}
$$

From the proof of Lemma 3, we have

$$A_3(z) - f_{X|Z}(x^*|z) \cdot \Pr(Z = z) = O(h_n) + O_P\Big(\frac{1}{\sqrt{nh_n}}\Big) \tag{B.8}$$

uniformly over z under Assumption 5 (ii), (iv) and (v), where $f_{X|Z}(x^*|z) \cdot \Pr(Z = z) \in (0, \infty)$ for all z by Assumption 5 (ii).

Similar lines of calculations yield

$$A_4(y, z) - F_{Y|X,Z}(y|x^*, z) \cdot f_{X|Z}(x^*|z) \cdot \Pr(Z = z) = O(h_n) + O_P\Big(\frac{1}{\sqrt{nh_n}}\Big) \tag{B.9}$$

uniformly over (y, z) under Assumption 5, where $F_{Y|X,Z}(y|x^*, z) \cdot f_{X|Z}(x^*|z) \cdot \Pr(Z = z)$ is uniformly bounded over (y, z) by Assumption 5 (ii)–(iii).

Conditionally on the event $|\widehat{x}^* - x^*| \le h_n$, we have

$$\begin{aligned} |\widehat{A}_3(z) - A_3(z)| \le & \frac{L_K \cdot |\widehat{x}^* - x^*|}{nh_n} \sum_{i=1}^{n} \mathbb{1}\{|X_i - x^*| \le 2h_n\} \cdot \mathbb{1}\{Z_i = z\} \\ = & 2 \cdot L_k \cdot |\widehat{x}^* - x^*| \left(f_{X|Z}(x^*|z) \cdot \Pr(Z = z) + o_P(1) \right) \\ = & O_P\left(|\widehat{x}^* - x^*| \right) \end{aligned}$$

uniformly over z under Assumption 5 (ii), (iv) and (v). Therefore, for any $\epsilon > 0$,

$$\begin{aligned} \Pr\left(\sup_{z \in \mathcal{Z}} |\widehat{A}_3(z) - A_3(z)| > \frac{\epsilon}{\sqrt{nh_n}} \right) = & \Pr\left(\sup_{z \in \mathcal{Z}} |\widehat{A}_3(z) - A_3(z)| > \frac{\epsilon}{\sqrt{nh_n}} \,\middle|\, |\widehat{x}^* - x^*| \le h_n \right) \cdot \Pr\left(|\widehat{x}^* - x^*| \le h_n \right) \\ & + \Pr\left(\sup_{z \in \mathcal{Z}} |\widehat{A}_3(z) - A_3(z)| > \frac{\epsilon}{\sqrt{nh_n}} \,\middle|\, |\widehat{x}^* - x^*| > h_n \right) \cdot \Pr\left(|\widehat{x}^* - x^*| > h_n \right) \end{aligned}$$

converges to zero due to $|\widehat{x}^* - x^*| = o_P\left(\frac{1}{nh_n} \wedge h_n\right)$ by Lemma 2 under Assumption 5 (i), (ii) and (iv). This shows that

$$\widehat{A}_3(z) - A_3(z) = o_P\left(\frac{1}{\sqrt{nh_n}}\right) \tag{B.10}$$

uniformly over z. Similar arguments show that

$$\widehat{A}_4(y, z) - A_4(y, z) = o_P\left(\frac{1}{\sqrt{nh_n}}\right) \tag{B.11}$$

uniformly over (y, z) under Assumption 5. Combining (B.7)–(B.11) together proves (B.1). A proof of (B.2) is analogous. □

B.6 Proof of Lemma 5

Proof. We can write

$$\widehat{\nu}_{\zeta,n}(y,\widehat{x}^*,z)-\tilde{\nu}_{\zeta,n}(y,\widehat{x}^*,x^*,z)=-\frac{1}{A_6(z)+(\widetilde{A}_6(z)-A_6(z))}\cdot(\widetilde{A}_5^{\zeta}(y,z)-\widehat{A}_5^{\zeta}(y,z))$$
$$-\frac{\widetilde{A}_6(z)-A_6(z)}{A_6(z)^2+A_6(z)\cdot(\widetilde{A}_6(z)-A_6(z))}\cdot\widehat{A}_2^{\zeta}(y,z)$$

where

$$\tilde{\nu}_{\zeta,n}(y,\widehat{x}^*,x^*,z)=\sum_{i=1}^{n}\zeta_i\frac{(\mathbb{1}\{Y_i\le y\}-F_{Y|X,Z}(y|X_i,z))\cdot K_{h_n}(X_i-\widehat{x}^*)\cdot\mathbb{1}\{Z_i=z\}}{\sqrt{nh_n}f_{X|Z}(x^*|z)\cdot Pr(Z=z)},$$

$$\widehat{A}_2^{\zeta}(y,z)=\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\zeta_i[\mathbb{1}\{Y_i\le y\}-F_{Y|X,Z}(y|X_i,z)]\cdot K_{h_n}(X_i-\widehat{x}^*)\cdot\mathbb{1}\{Z_i=z\},$$
$$\widehat{A}_5^{\zeta}(y,z)=\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\zeta_iF_{Y|X,Z}(y|X_i,z)\cdot K_{h_n}(X_i-\widehat{x}^*)\cdot\mathbb{1}\{Z_i=z\},$$
$$A_6(z)=f_{X|Z}(x^*|z)\cdot\Pr(Z=z),$$

and

$$\widetilde{A}_5^{\zeta}(y,z)=\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\zeta_i\widehat{F}_{Y|X,Z}(y|X_i,z)\cdot K_{h_n}(X_i-\widehat{x}^*)\cdot\mathbb{1}\{Z_i=z\},$$
$$\widetilde{A}_6(z)=\widehat{f}_{X|Z}(\widehat{x}^*|z)\cdot\widehat{\Pr}(Z=z).$$

First,

$$|\widetilde{A}_5^{\zeta}(y,z)-\widehat{A}_5^{\zeta}(y,z)|\le\left(\sup_{(y,x,z)}|\widehat{F}_{Y|X,Z}(y|x,z)-F_{Y|X,Z}(y|x,z)|\right)\cdot\left(|A_7^{\zeta}(z)|+|\widehat{A}_7^{\zeta}(z)-A_7^{\zeta}(z)|\right),$$

where

$$\widehat{A}_7^{\zeta}(z)=\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\zeta_iK_{h_n}(X_i-x^*)\cdot\mathbb{1}\{Z_i=z\}$$
$$A_7^{\zeta}(z)=\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\zeta_iK_{h_n}(X_i-\widehat{x}^*)\cdot\mathbb{1}\{Z_i=z\}$$

Note that $\sup_{(y,x,z)}|\widehat{F}_{Y|X,Z}(y|x,z)-F_{Y|X,Z}(y|x,z)|=o_P(1)$ under Assumption 5 (iii), (iv) and (v). Under Assumption 5 (ii), (iv) and (v), we have $A_7^{\zeta}(z)=O_P(1)$ uniformly over z, similarly to the argument that shows $A_2(y,z) = O_P(1)$ uniformly over (y,z) in the proof of Lemma 3. Also, under Assumption 5 (ii), (iv) and (v), we have $\widehat{A}_7^{\zeta}(z) - A_7^{\zeta}(z) = o_P(1)$ uniformly over z, similarly to the proof of Lemma 4. It therefore follows that

$$\widetilde{A}_5^{\zeta}(y,z) - \widehat{A}_5^{\zeta}(y,z) = o_P(1)$$

uniformly over (y,z).

Second, we have $\widetilde{A}_2^{\zeta}(y,z) = O_P(1)$ uniformly over (y,z), similarly to the argument that shows $A_2(y,z) = O_P(1)$ uniformly over (y,z) in the proof of Lemma 3.

Finally, $\widetilde{A}_6(z) - A_6(z) = o_P(1)$ uniformly over z under Assumption 5 (ii), (iv) and (v) by similar arguments to the proof of Lemma 4, and $A_6(z)$ bounded away from zero and infinity uniformly over z by Assumption 5 (ii).

Putting all the above arguments together, we have

$$\widehat{\nu}_{\zeta,n}(y,\widehat{x}^*,z) - \tilde{\nu}_{\zeta,n}(y,\widehat{x}^*,x^*,z) = o_P(1)$$

uniformly over (y,z). □

C Proofs on the generalized location-scale models

C.1 Characterization of Assumption 4

To see the sufficiency, note that in the first case, we simply have $F_{X|Z}(\cdot|z') = F_{X|Z}(\cdot|z)$. In the second case, let $(z,z') \in \mathcal{Z}^2$ be such that $\sigma(z') \neq \sigma(z)$. Then one can check that $F_{X|Z}(x^*(z,z')|z) = F_{X|Z}(x^*(z,z')|z')$ with

$$x^*(z,z') = \mu\left[\frac{\sigma(z)\psi(z') - \sigma(z')\psi(z)}{\sigma(z) - \sigma(z')}\right]. \tag{C.1}$$

Hence, Assumption 4 holds. On the other hand, if $\sigma(\cdot)$ is constant and ψ is one-to-one, then, for any $z \neq z'$, we either have $F_{X|Z}(x|z) > F_{X|Z}(x|z')$ for all $x \in \text{Int}(\text{Supp}(X|Z=z))$ or $F_{X|Z}(x|z) < F_{X|Z}(x|z')$ for all $x \in$Int(Supp$(X|Z=z)$). □

C.2 Verification of Assumption 8

If $\sigma(\cdot)$ is not constant, let $(z, z') \in \mathcal{Z}^2$ be such that $\sigma(z') \neq \sigma(z)$. Then $F_{X|Z}(x|z) = F_{X|Z}(x|z')$ if and only if $x = x^*(z, z')$ defined in (C.1). Thus Assumption 8 holds. If $\sigma(\cdot)$ is constant but $\psi(\cdot)$ is not, then we either have $F_{X|Z}(x|z) > F_{X|Z}(x|z')$ for all x or $F_{X|Z}(x|z) < F_{X|Z}(x|z')$, so Assumption 8 holds as well. □

C.3 Verification of Assumptions 4 and 12

We already showed above that Assumption 4 holds. Next, Assumption 12(i) holds if $\beta_0 = 0$. So let us suppose hereafter that $\beta_0 \neq 0$. Given (3.1) and $\mu(x) = x$, we have

$$q_{zz'}(x) = \psi(z') + \frac{\sigma(z')}{\sigma(z)}(x - \psi(z)). \tag{C.2}$$

For any (z_1, z_1') such that $\sigma(z_1) \neq \sigma(z_1')$, let us define x_1 by

$$x_1 = \frac{1}{\sigma(z_1) - \sigma(z_1')} \left[\sigma(z_1)\psi(z_1') - \sigma(z_1')\psi(z_1) + \sigma(z_1)(z_1' - z_1)\beta_0/\alpha_0\right]. \tag{C.3}$$

Then some calculations show that

$$x_1\alpha_0 + z_1\beta_0 = q_{z_1z_1'}(x_1)\alpha_0 + z_1'\beta_0. \tag{C.4}$$

Now, recall that

$$\begin{aligned} E(Y|X = x_1, Z = z_1) =& E[g_1(x_1\alpha_0 + z_1\beta_0, \varepsilon)|\eta = F_{X|Z}(x_1|z_1)], \\ E(Y|X = q_{z_1z_1'}(x_1), Z = z_1') =& E[g_1(q_{z_1z_1'}(x_1)\alpha_0 + z_1'\beta_0, \varepsilon)|\eta = F_{X|Z}(x_1|z_1)]. \end{aligned}$$

Thus, (C.4) implies that

$$E[Y|X = x_1, Z = z_1] = E[Y|X = q_{z_1z_1'}(x_1), Z = z_1'],$$

and Assumption 12 holds. □

C.4 Verification of Assumptions 16-17

We first show that Assumption 16 holds as long as $\sigma(\cdot)$ is not constant. Take any $z_0 \in \mathcal{Z}$. For any $z \in \mathcal{Z}$, we either have $\sigma(z) \neq \sigma(z_0)$ or $\sigma(z) = \sigma(z_0)$. In the first case, $x^*(z, z_0)$ defined by (C.1) satisfies $F_{X|Z}(x^*(z, z_0)|z) = F_{X|Z}(x^*(z, z_0)|z_0) \in (0, 1)$ and thus z is connected to z_0. In the second case, because $\sigma(\cdot)$ is not constant, there exists z' such that $\sigma(z') \neq \sigma(z) = \sigma(z_0)$. But then, we have $F_{X|Z}(x^*(z, z')|z) = F_{X|Z}(x^*(z, z')|z') \in (0, 1)$ and $F_{X|Z}(x^*(z_0, z')|z_0) = F_{X|Z}(x^*(z_0, z')|z') \in (0, 1)$. So again, z is connected to z_0.

Next, we show that Assumption 17 holds if 1, $\psi(\cdot)$ and $\sigma(\cdot)$ are not linearly related on any interval. Let us assume that Assumption 17 does not hold. Then there exists $u \in \mathbb{R}$ and an interval I on which $h(\cdot, u)$ is constant. Because μ is one-to-one, this means that $z \mapsto \psi(z) + \sigma(z)u$ is constant on I. But this contradicts the fact that 1, $\psi(\cdot)$ and $\sigma(\cdot)$ are linearly independent on I. $\square$