Identification of Instrumental Variable Correlated Random Coefficients Models*

Matthew A. Masten† Alexander Torgovitsky‡

January 18, 2016

Abstract

We study identification and estimation of the average partial effect in an instrumental variable correlated random coefficients model with continuously distributed endogenous regressors. This model allows treatment effects to be correlated with the level of treatment. The main result shows that the average partial

*This paper was presented at the UW–Milwaukee Mini-Conference on Microeconometrics, Duke, Northwestern, Chicago, the Harvard/MIT joint workshop, Ohio State, Boston College, the Triangle econometrics conference, University College London, the First Conference in Econometric Theory at Universidad de San Andrés, and the 2014 Cowles Summer Conference on Econometrics. We thank audiences at those seminars as well as Federico Bugni, Ivan Canay, Matias Cattaneo, Andrew Chesher, Rob Garlick, Bryan Graham, Jim Heckman, Stefan Hoderlein, Joel Horowitz, Max Kasy, Shakeeb Khan, Jia Li, Simon Lee, Arnaud Maurel, Whitney Newey, Maya Rossin-Slater, Christoph Rothe, Elie Tamer, Duncan Thomas, Chris Timmins, and Ed Vytlacil for helpful conversations and comments. We thank the editor and the referees for their suggestions. We thank Fu Ouyang for research assistance.

†Department of Economics, Duke University, matt.masten@duke.edu

‡Department of Economics, Northwestern University, a-torgovitsky@northwestern.edu

effect is identified by averaging coefficients obtained from a collection of ordinary linear regressions that condition on different realizations of a control function. These control functions can be constructed from binary or discrete instruments which may affect the endogenous variables heterogeneously. Our results suggest a simple estimator that can be implemented with a companion Stata module.

JEL classification: C14; C26; C51

Keywords: correlated random coefficients, instrumental variables, unobserved heterogeneity, semiparametrics

1. Introduction

This paper is about the linear correlated random coefficients (CRC) model. In its simplest form, the model can be written as

$$Y = B_0 + B_1 X, \tag{1}$$

where Y is a continuous outcome variable, X is an endogenous explanatory variable, and $B \equiv (B_0, B_1)$ are unobservable variables that may be statistically dependent with X. Endogeneity in X is often addressed by using the variation of an instrumental variable, Z, that is plausibly independent (or uncorrelated) with (B_0, B_1), but correlated with X. The most common estimator for implementing this strategy is two-stage least squares (2SLS).

If both X and Z are binary, and if Z affects X monotonically, then the 2SLS estimator is consistent for the local average treatment effect (LATE), which is the unweighted average of B_1 for the subpopulation of compliers (Imbens and Angrist 1994). However, as the support of X grows from binary to multi-valued discrete to continuous, the 2SLS estimand becomes an increasingly complicated weighted average of LATEs between different X realizations (Angrist and Imbens 1995, Angrist, Graddy, and Imbens 2000). This type of weighted average of LATEs can be difficult to interpret. These difficulties compound as additional explanatory variables are added to (1) (Abadie 2003).

A more natural parameter is $\mathbb{E}[B_1]$, which is often called the average partial effect (APE). In a series of papers, Heckman and Vytlacil (1998) and Wooldridge (1997, 2003, 2008) showed that if the causal effect of Z on X is homogeneous, then the 2SLS estimator is consistent for the APE. This modeling assumption is uncomfortably asymmetric: treatment effects may be heterogeneous, but instrument effects may not. When the causal effect of the instrument is actually heterogeneous, the 2SLS estimand generally differs from the APE.

As an alternative to assuming homogeneity in the first stage, one can consider different instrumental variables estimators besides 2SLS. Florens, Heckman, Meghir, and Vytlacil (2008) take this approach in considering a polynomial version of (1) that also includes an additive nonparametric function of X. Like Imbens and Newey (2009) and others, they assume that X is continuous, that there exists a function h that is strictly increasing in a scalar unobservable V such that $X = h(Z, V)$, and that Z is independent with the unobservables in both the outcome and first stage equations. These restrictions still allow for heterogeneity in the causal effect of Z on X, albeit in a limited form. Under these assumptions, Florens et al. (2008) show that the APE can be identified if Z is continuous.

We show that for the linear CRC model (1), this same first stage assumption can be sufficient to identify the APE even if Z has binary or discrete support, which is often the case for instruments used in practice. Moreover, our main identification result suggests a simple semiparametric estimator of the APE that does not suffer from the curse of dimensionality. A companion Stata module for implementing the estimator is available. Taken together, our results provide an alternative to 2SLS for applied researchers studying the causal effect of a continuous treatment X using a binary or discrete instrument Z. The benefit of our estimator relative to 2SLS is that it is consistent for an easily-interpretable parameter while still allowing for discrete instruments and heterogeneity in the causal effect of Z on X.

2. Model and Identification

A more general version of (1) is

$$Y = B_0 + \sum_{j=1}^{d_x} B_j X_j + \sum_{j=1}^{d_1} B_{d_x+j} Z_{1j} \equiv W'B, \tag{2}$$

where $X \in \mathbb{R}^{d_x}$ is a vector of potentially endogenous variables, $Z_1 \in \mathbb{R}^{d_1}$ is a vector of included exogenous variables with j^{th} component Z_{1j}, $W \equiv [1, X', Z_1']' \in \mathbb{R}^{d_w}$ with $d_w \equiv 1 + d_x + d_1$, and $B \in \mathbb{R}^{d_w}$ is a vector of unobservable variables. In addition to Z_1, there is a vector of excluded exogenous variables (instruments) $Z_2 \in \mathbb{R}^{d_2}$ that do not directly affect Y in (2). We write the exogenous variables together as $Z \equiv [Z_1', Z_2']' \in \mathbb{R}^{d_z}$ with $d_z \equiv d_1 + d_2$.

We divide the vector of endogenous variables X into subvectors of lengths $d_b \geq 1$ and $d_x - d_b \geq 0$. We call the first d_b components of X the *basic* endogenous variables and the last $d_x - d_b$ components of X the *derived* endogenous variables. We assume that the basic endogenous variables satisfy a particular first stage structure specified ahead. In contrast, we assume that the derived endogenous variables are known functions of the basic endogenous variables and the included exogenous variables Z_1. For example, we could have $d_b = 1$ and derived endogenous variables $X_k = X^k$ for $k > d_b$, as in Florens et al. (2008). Or, we could have an interaction term $X_k = X_1 Z_1$ for some $k > d_b$, thereby allowing for the distribution of causal effects of X_1 on Y to differ arbitrarily across values of Z_1. For example, this allows men and women to have different distributions of treatment effects, allowing for heterogeneity on observables to be dealt with in the usual way.

Throughout our analysis we frequently use the observable random d_b vector

$$R \equiv [\, R_1, \, \ldots, \, R_{d_b} \,]' \equiv [\, F_{X_1|Z}(X_1|Z), \, \ldots, \, F_{X_{d_b}|Z}(X_{d_b}|Z) \,]',$$

which we call the *conditional rank* of X. Below, we restrict X_k to be continuously distributed for $k = 1, \ldots, d_b$ so that each R_k is distributed uniformly on $[0, 1]$. Our main result is theorem

1, which uses the following assumptions.

Assumption I.

I1. (Existence of moments) $\mathbb{E}[B]$ *and* $\mathbb{E}[WW']$ *exist.*

I2. (First stage equation) *For each basic endogenous variable* X_k, $k = 1, \ldots, d_b$, *there exists a scalar unobservable random variable* V_k *and a possibly unknown function* h_k *that is strictly increasing in its second argument, for which* $X_k = h_k(Z, V_k)$. *The vector* $V \equiv (V_1, \ldots, V_{d_b})$ *is continuously distributed.*

I3. (Derived endogenous variables) *For each* $k = d_b + 1, \ldots, d_x$, *there exists a known function* g_k *such that* $X_k = g_k(X_1 \ldots, X_{d_b}, Z_1)$.

I4. (Instrument exogeneity) $(B, V) \perp\!\!\!\perp Z$.[1]

I5. (Instrument relevance) $\mathbb{E}[WW'|R = r]$ *is invertible for almost every* r *in a known Lebesgue measurable set* $\mathcal{R} \subseteq \text{supp}(R)$.

Theorem 1. *Define* $\beta(r) \equiv \mathbb{E}[B|R = r]$ *and* $\widetilde{\beta}(r) \equiv \mathbb{E}[WW'|R = r]^{-1}\,\mathbb{E}[WY|R = r]$. *Under Assumptions I,* $\widetilde{\beta}(r) = \beta(r)$ *for any* $r \equiv (r_1, \ldots, r_{d_b}) \in \mathcal{R}$. *Hence both* $\beta(r)$ *and* $\beta_{\mathcal{R}} \equiv \mathbb{E}[B|R \in \mathcal{R}]$ *are point identified.*

The proof of theorem 1 uses the following implication of I2–I4. Similar results have been used previously by Imbens (2007), Florens et al. (2008), Imbens and Newey (2009), and Torgovitsky (2015). Our working paper (Masten and Torgovitsky 2014) contains a proof.

Proposition 1. *I2 and I4 imply that* $(R, B) \perp\!\!\!\perp Z$. *If I3 also holds, then* $W \perp\!\!\!\perp B|R$.

[1] Our main results, theorems 1 and 2, continue to hold under the weaker assumption that $V \perp\!\!\!\perp Z$ and $\mathbb{E}[B|V, Z] = \mathbb{E}[B|V]$, although several of our discussion points use the stronger assumption I4.

Proof of theorem 1. I1 ensures that all conditional moments of interest exist. Premultiplying both sides of (2) by W and taking expectations conditional on $R = r$, we have

$$\mathbb{E}[WY|R = r] = \mathbb{E}[WW'B|R = r] = \mathbb{E}[WW'|R = r]\beta(r), \tag{3}$$

by proposition 1. Given I5, we can premultiply both sides of (3) by the inverse of $\mathbb{E}[WW'|R = r]$ to obtain $\widetilde{\beta}(r) = \beta(r)$ for any $r \in \mathcal{R}$. Since $\widetilde{\beta}(r)$ is a feature of the distribution of observables, this shows that $\beta(r)$ is identified. Because the distribution of R is observable and $\mathcal{R}$ is known, $\beta_{\mathcal{R}}$ is also identified. *Q.E.D.*

3. Discussion

The intuition behind theorem 1 is that the first stage (I2) and instrument exogeneity (I4) assumptions imply that $R_k = F_{V_k}(V_k)$ for $k = 1, \ldots, d_b$, so that conditioning on $R = r$ is the same as conditioning on V_k being equal to its r_kth quantile $Q_{V_k}(r_k)$ for $k = 1, \ldots, d_b$. Hence, conditioning on R fixes V, so that all of the remaining variation in the basic endogenous variables is driven solely by variation in Z. Since the derived endogenous variables are known functions of the basic endogenous variables and Z_1, all of the variation in W conditional on $R = r$ is also due solely to variation in Z. Given that conditioning on $R = r$ is equivalent to conditioning on $V_k = Q_{V_k}(r_k)$ for all $k = 1, \ldots, d_b$, this together with instrument exogeneity (I4) then implies that B is independent of W, conditional on $R = r$. As a result, a linear regression of Y on W conditional on $R = r$ identifies $\beta(r) \equiv \mathbb{E}[B|R = r]$. As in standard linear regression models, this requires $\mathbb{E}[WW'|R = r]^{-1}$ to exist, which is assumed in I5. Averaging $\mathbb{E}[B|R = r]$ over $r \in \mathcal{R}$ then yields $\beta_{\mathcal{R}} \equiv \mathbb{E}[B|R \in \mathcal{R}]$. If I5 holds for some measure one subset of $\text{supp}(R)$, then $\beta_{\mathcal{R}} = \mathbb{E}[B]$ and the APE is identified.

Assumption I5 serves the same purpose as the standard no perfect multicollinearity condition in ordinary least squares. Consequently, it requires the analyst to avoid standard causes of failure, such as the dummy variable trap. Aside from these mechanical causes of failure, whether I5 holds depends on the relevance of the instruments. When $d_x = d_b = 1$, so that there is a single basic endogenous variable and no derived endogenous variables, I5 is equivalent to $\text{Var}[Q_{X|Z}(r \mid Z)] > 0$ for all $r \in \mathcal{R}$. If $Z \in \{0, 1\}$ is binary and nondegenerate, then $\text{Var}[Q_{X|Z}(r \mid Z)] > 0$ is true if and only if $Q_{X|Z}(r \mid 0) \neq Q_{X|Z}(r \mid 1)$. That is, the two curves in figure 1 must be separated at r. Since $Q_{X|Z}(r \mid Z) = h(Z, Q_V(r))$ by I2 and I4, I5 requires that for each $r \in \mathcal{R}$ there are distinct $z, z' \in \text{supp}(Z)$ with $h(z, Q_V(r)) \neq h(z', Q_V(r))$. Hence, the instrument must affect the endogenous variable for all units with first stage unobservables $v = Q_V(r)$ at which we want to learn the conditional mean of B.

Whether I5 holds is an empirical matter in the sense that the condition only depends on the distribution of observables and so, at least in principle, can be checked in the data. If I5 holds for a subset $\mathcal{R}$ of $\text{supp}(R)$, then theorem 1 identifies $\beta_{\mathcal{R}} \equiv \mathbb{E}[B|R \in \mathcal{R}]$, which may not equal $\mathbb{E}[B]$. Nevertheless, $\beta_{\mathcal{R}}$ has an interpretation similar to the unweighted LATE of Imbens and Angrist (1994). That is, $\beta_{\mathcal{R}}$ is the unweighted average of B for those agents for whom the instrument Z has a causal effect on their treatment X. Unlike Imbens and Angrist (1994), we do not require this effect to be monotonic. This type of parameter may be of comparable (or even greater) interest than $\mathbb{E}[B]$ for a policy maker considering a policy change that affects the determination of X through an incentive Z. While I5 may fail for some subset of $\text{supp}(R)$, it is an intuitively appealing requirement for an instrument. Agents in a subpopulation $R = r$ for an r at which $\mathbb{E}[WW'|R = r]$ is singular do not experience independent variation in W due to variation in Z, and so it is natural that $\mathbb{E}[B|R = r]$ should not be point identified for those agents.

If a component of B is bounded, then knowledge of $\mathbb{E}[B|R \in \mathcal{R}]$ for a subset $\mathcal{R}$ of $\text{supp}(R)$ yields partial identification of $\mathbb{E}[B]$. For example, consider the model (1) and suppose that $B_1 \in [B_{1L}, B_{1U}]$, where B_{1L} and B_{1U} are known constants specified by the researcher. Then bounds on $\mathbb{E}[B_1]$ follow from the law of iterated expectation similar to Manski (1989):

$$\mathbb{E}[B_1|R \in \mathcal{R}]\mathbb{P}[R \in \mathcal{R}] + B_{1L}\mathbb{P}[R \notin \mathcal{R}] \leq \mathbb{E}[B_1] \leq \mathbb{E}[B_1|R \in \mathcal{R}]\mathbb{P}[R \in \mathcal{R}] + B_{1U}\mathbb{P}[R \notin \mathcal{R}].$$

As in standard linear models, identification of $\mathbb{E}[B]$ (or a conditional on $\mathcal{R}$ version) via theorem 1 provides identification of the APE when the outcome equation includes nonlinear functions of X or interactions with covariates Z_1. This is an elementary point, but we mention it for clarity. Suppose for example that $Y = B_0 + B_1 X + B_2 X Z_1$. Then the APE is given by $\mathbb{E}[B_1] + \mathbb{E}[B_2]\,\mathbb{E}[Z_1]$, which can be obtained from estimates of $\mathbb{E}[B_1]$, $\mathbb{E}[B_2]$ and $\mathbb{E}[Z_1]$. Alternatively, an analyst may be interested in the APE for some predetermined value of z_1, which would be given by $\mathbb{E}[B_1] + \mathbb{E}[B_2]z_1$. If (2) contains nonlinear terms, e.g. $Y = B_0 + B_1 X + B_2 X^2$, then an analyst may be more interested in reporting $\mathbb{E}[B_1] + 2\,\mathbb{E}[B_2]x$ as the APE when X is exogenously set to x. These quantities can all be obtained after applying theorem 1.

If (2) is misspecified and instead $Y = g(W, B)$ for some nonlinear function g, then standard arguments imply that $\widetilde{\beta}(r)$ is the minimal mean-squared error linear approximation to $\mathbb{E}[Y|W = w, R = r] = \mathbb{E}[g(w, B)|W = w, R = r] = \mathbb{E}[g(w, B)|R = r]$, using proposition 1. This latter quantity is what Blundell and Powell (2003) call the average structural function at $W = w$, but conditional on $R = r$. The implication is that an otherwise consistent estimator of $\beta_{\mathcal{R}}$ will, under misspecification, still be consistent for an average of best linear approximations to the average structural function, conditional on $R \in \mathcal{R}$.

To see the benefit of Theorem 1 relative to 2SLS, suppose that I4 holds for $Z \in \{0, 1\}$, and write $X = h(0, V) + \Delta \cdot Z$, where $\Delta = h(1, V) - h(0, V)$ is a random variable representing the distribution of instrument effects. Then the 2SLS estimand equals $\beta_{\text{2SLS}} = \mathbb{E}(B_1\Delta)/\,\mathbb{E}(\Delta)$, so that $\beta_{\text{2SLS}} - \mathbb{E}(B_1) = \text{Cov}(B_1, \Delta)/\,\mathbb{E}(\Delta)$. This bias depends on the direction and magnitude of linear dependence between the treatment effect B_1 and the instrument effect Δ. It also depends on the magnitude of the average causal effect of Z on X, $\mathbb{E}(\Delta)$. The bias is generally non-zero if Δ is non-degenerate, in which case the 2SLS estimator is not consistent for the APE.

Our analysis is related to work by Jun (2009), who studies a linear random coefficients version of Chesher's (2003) nonparametric model. Both papers maintain a first stage assumption similar to I2. Jun also observes that the linearity of his outcome equation allows for discrete instruments. However, Jun requires the coefficients in the outcome equation to all be determined by a single scalar unobservable, conditional on V. This type of co-monotonicity restricts the dimension of unobserved heterogeneity in the outcome equation to 1, conditional on V. In contrast, our outcome equation still allows for high dimensional heterogeneity after conditioning on V. In particular, the joint distribution of our random coefficients B can have full support on $\mathbb{R}^{d_w}$.

Theorem 1 complements a result by Florens et al. (2008). Those authors consider a model with a single basic endogenous variable X and the outcome equation

$$Y = \varphi(X) + B_0 + B_1 X + B_2 X^2 + \cdots + B_K X^K,$$

for some known K, where φ is an unknown function, and $(B_0, \ldots, B_K)$ are random coefficients that may be dependent with X. Except for φ, this outcome equation can be obtained from (2) with basic endogenous variable X, and derived endogenous variables $(X^2, \ldots, X^K)$. The cost of including the φ function is that Florens et al. (2008) require a continuous instrument—see the step from equation 10 to the next line on page 1203. We do not include the φ function, but are generally able to achieve identification of the average coefficients in the polynomial outcome equation model so long as the distribution of Z has at least $K + 1$ support points. Florens et al. (2008) also maintain I2 and I4, but in place of I5 they impose a "measurable separability" condition that is somewhat high-level. As those authors discuss, their measurable separability condition may fail if the first stage equation is not continuous in V. Theorem 1 does not require such continuity. This also allows for the support of X conditional on Z to be disjoint.

Among the maintained assumptions for theorem 1, I2 is generally the most controversial. While it is more general than the specifications of Heckman and Vytlacil (1998) and Wooldridge (1997, 2003, 2008), which impose a homogeneous causal effect of Z on X, it does restrict the basic endogenous variables to be continuous and limits the unobserved heterogeneity in their first stage equations to have dimension one. One-dimensional unobservable heterogeneity of the sort in I2 can be interpreted as "rank invariance" (Doksum 1974) in the causal effect of Z on each basic component of X. Rank invariance means that the ordinal ranking of any two agents in terms of any component of X_k ($k \leq d_b$) would be the same if both agents received the same realization of Z, for any realization of Z. See Heckman, Smith, and Clements (1997), Chernozhukov and Hansen (2005) and Torgovitsky (2015) for further discussions of rank invariance. While one-dimensional heterogeneity is restrictive, there are few alternatives in the literature that allow for high-dimensional heterogeneity in both the outcome and first stage equations while still attaining point identification of a broadly interpretable parameter. One such example is Masten (2012), who allows for linear random coefficients in both the outcome and first stage equations. His results require a continuous instrument with small support.

In addition to $\mathbb{E}[B]$, we can also identify what Florens et al. (2008) refer to as the "average effect of the treatment on the treated."

Theorem 2. *Let* $\widetilde{X} \equiv (X_1, \ldots, X_{d_b})$. *Under Assumptions I, the "average effect of the treatment on the treated,"* $\mathbb{E}[B \mid \widetilde{X} = x]$, *is point identified for any* $x \in \operatorname{supp}(\widetilde{X})$ *such that*

$$\left\{ \left(F_{X_1|Z}(x_1|z), \ldots, F_{X_{d_b}|Z}(x_{d_b}|z) \right) : z \in \operatorname{supp}(Z|\widetilde{X} = x) \right\} \subseteq \mathcal{R}. \tag{4}$$

Proof of theorem 2. From the proof of theorem 1, $\beta(r) \equiv \mathbb{E}[B \mid R = r]$ is identified for all $r \in \mathcal{R}$. By iterated expectations, the definition of R, and proposition 1, we have

$$\mathbb{E}[B \mid \widetilde{X} = x] = \mathbb{E}_{R|\widetilde{X}}\left[\mathbb{E}\left(B \mid \widetilde{X} = x, R\right) \mid \widetilde{X} = x\right]$$

$$= \mathbb{E}_{Z|\widetilde{X}}\left[\mathbb{E}\left(B \mid \widetilde{X} = x, R = (F_{X_1|Z}(x_1|Z), \ldots, F_{X_{d_b}|Z}(x_{d_b}|Z))\right) \mid \widetilde{X} = x\right]$$

$$= \mathbb{E}_{Z|\widetilde{X}}\left[\mathbb{E}\left(B \mid R = (F_{X_1|Z}(x_1|Z), \ldots, F_{X_{d_b}|Z}(x_{d_b}|Z))\right) \mid \widetilde{X} = x\right]$$

$$= \mathbb{E}_{Z|\widetilde{X}}\left[\beta\left((F_{X_1|Z}(x_1|Z), \ldots, F_{X_{d_b}|Z}(x_{d_b}|Z))\right) \mid \widetilde{X} = x\right],$$

which is identified since $(F_{X_1|Z}(x_1|z), \ldots, F_{X_{d_b}|Z}(x_{d_b}|z)) \in \mathcal{R}$ for all $z \in \text{supp}(Z|\widetilde{X} = x)$. *Q.E.D.*

Condition (4) in theorem 2 holds trivially if $\mathcal{R} = [0,1]^{d_b}$. To interpret (4) when $\mathcal{R}$ is a strict subset of $[0,1]^{d_b}$, suppose for simplicity that $d_b = 1$. Then (4) requires that if $x = Q_{X|Z}(r|z)$ for some $z \in \text{supp}(Z|\widetilde{X} = x)$, then $r \in \mathcal{R}$, so that $\beta(r)$ is identified. In the simple model (1), a sufficient condition for (4) is $\text{Var}[F_{X|Z}(x|Z)] > 0$. To see this, suppose z and z' are such that $r \equiv F_{X|Z}(x|z) > F_{X|Z}(x|z') \equiv r'$. Then $Q_{X|Z}(r|z') > x \geq Q_{X|Z}(r|z)$ and hence $r \in \mathcal{R}$. The strict inequality follows by $F_{X|Z}(x|z') < r$ since X is continuous. The weak inequality follows by $r = F_{X|Z}(x|z)$ and the definition of the quantile. A symmetric argument shows that $r' \in \mathcal{R}$ as well. For example, in figure 1, $x_0 = Q_{X|Z}(.8|0) = Q_{X|Z}(.6|1)$ with $\beta(r)$ identified at both $r = 0.6$ and $r' = 0.8$. Hence we have that $\mathbb{E}(B_1|X = x_0) = \beta(0.6)\,\mathbb{P}(Z = 1|X = x_0) + \beta(0.8)\,\mathbb{P}(Z = 0|X = x_0)$ is point identified.

The average effect of the treatment on the treated provides one way of exploring heterogeneity in treatment effects. A truly constant treatment effect would yield a function $\mathbb{E}[B \mid \widetilde{X} = x]$ which is constant over x. An increasing function would indicate positive correlation between received treatment and the coefficients, while a decreasing function would indicate negative correlation between received treatment and the coefficients. Indeed, if $\mathcal{R} = [0,1]^{d_b}$ then $\mathbb{E}[B \mid \widetilde{X} = x]$ is identified for all $x \in \text{supp}(\widetilde{X})$ and hence the correlations $\mathbb{E}[B_j X_k] = \mathbb{E}[\mathbb{E}(B_j \mid \widetilde{X}) X_k]$ are also identified for $k = 1, \ldots, d_b$ and any j.

It is also possible to identify moments of the distribution of B other than the mean. For example, (1) implies that $Y^2 = B_0^2 + 2B_0 B_1 X + B_1^2 X^2$. If Z has three or more points of support, then applying theorem 1 to Y^2 provides identification of $\mathbb{E}[B_0^2 | R \in \mathcal{R}]$, $\mathbb{E}[B_1^2 | R \in \mathcal{R}]$ and $\mathbb{E}[B_0 B_1 | R \in \mathcal{R}]$. Combining these parameters with $\mathbb{E}[B_0 | R \in \mathcal{R}]$ and $\mathbb{E}[B_1 | R \in \mathcal{R}]$ provides identification of all conditional second moments of the vector $B = (B_0, B_1)'$. In principle, this argument could be continued to identify higher moments of B as well.

4. Estimation

We conclude by briefly sketching a sample analog estimator of the APE based on theorem 1. Our working paper (Masten and Torgovitsky 2014) contains more details on implementation and asymptotic theory. The estimator is essentially an average of ordinary linear regressions, each run conditional on a realization of a control function, and so it shares similarities with the control function approaches of, for example, Heckman and Robb (1985), Blundell and Powell (2004), Imbens and Newey (2009), Rothe (2009), Jun (2009), Torgovitsky (2013), and Hoderlein and Sherman (2015). For simplicity, suppose there is only one basic endogenous variable ($d_b = 1$) denoted by X, although there may be any number of known exogenous and derived endogenous variables Z. Let $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ be an i.i.d. sample of (Y, X, Z).

There are three steps to computing the estimator. First, construct estimates $\widehat{R}_i$ of $R_i \equiv F_{X|Z}(X_i | Z_i)$ by replacing $F_{X|Z}$ with a consistent estimator. If Z has high dimension, this can be done by modeling $Q_{X|Z}$ through linear quantile regression and inverting—see our working paper for more details. Second, estimate $\beta(r) \equiv \mathbb{E}[B \mid R = r]$ by a smoothed sample analog of its expression in theorem 1, that is,

$$\widehat{\beta}(r) \equiv \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{k}_i^h(r) W_i W_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{k}_i^h(r) W_i Y_i \right),$$

where $\widehat{k}_i^h(r) \equiv h^{-1}K((\widehat{R}_i - r)/h)$ are weights constructed through an ordinary second-order kernel function K with bandwidth parameter h that tends to 0 asymptotically. Third, estimate $\beta_{\mathcal{R}} \equiv \mathbb{E}[B \mid R \in \mathcal{R}]$ by averaging the second stage estimator, $\widehat{\beta}_{\mathcal{R}} \equiv \lambda(\mathcal{R})^{-1} \int_{\mathcal{R}} \widehat{\beta}(r)\, dr$, where $\lambda(\mathcal{R})$ is the Lebesgue measure of $\mathcal{R} \subseteq [0,1]$, e.g., $\lambda(\mathcal{R}) = .4$ if $\mathcal{R} = [0.2, 0.6]$. In our working paper, we establish relatively weak low-level conditions under which $\sqrt{n}(\widehat{\beta}_{\mathcal{R}} - \beta_{\mathcal{R}})$ is asymptotically normal. Its limiting variance is complicated, but the percentile bootstrap can be used for inference. A Stata module for implementing this estimator is available on our websites.

References

Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263, April 2003. ISSN 0304-4076. URL http://www.sciencedirect.com/science/article/pii/S0304407602002014. 3

Joshua D. Angrist and Guido W. Imbens. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *The Journal of the American Statistical Association*, 90(430):431–442, 1995. 2

Joshua D Angrist, Kathryn Graddy, and Guido W. Imbens. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies*, 67(3):499–527, 2000. 2

Richard W Blundell and James L Powell. Endogeneity in Nonparametric and Semiparametric Regression Models. In Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, editors, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, pages 312–357, Cambridge, United Kingdom, 2003. Cambridge University Press. 8

Richard W. Blundell and James L. Powell. Endogeneity in Semiparametric Binary Response

Models. *The Review of Economic Studies*, 71(3):655–679, July 2004. ISSN 00346527. URL http://www.jstor.org/stable/3700739. 12

Victor Chernozhukov and Christian B Hansen. An IV Model of Quantile Treatment Effects. *Econometrica*, 73(1):245–261, 2005. 10

Andrew Chesher. Identification in Nonseparable Models. *Econometrica*, 71(5):1405–1441, 2003. 8

Kjell Doksum. Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case. *The Annals of Statistics*, 2(2):267–277, 1974. 10

Jean-Pierre Florens, James J Heckman, Costas Meghir, and Edward Vytlacil. Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects. *Econometrica*, 76(5):1191–1206, 2008. 3, 4, 5, 9, 10

James J Heckman and Richard Robb. Alternative Methods for Evaluating the Impact of Interventions: An Overview. *Journal of Econometrics*, 30(1):239–267, 1985. 12

James J Heckman and Edward Vytlacil. Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling. *The Journal of Human Resources*, 33(4):974–987, 1998. 3, 9

James J. Heckman, Jeffrey Smith, and Nancy Clements. Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies*, 64(4):487–535, 1997. 10

Stefan Hoderlein and Robert Sherman. Identification and Estimation in a Correlated Random Coefficients Binary Response Model. *Journal of Econometrics*, 188(1):135–149, 2015. 12

Guido W. Imbens. Nonadditive Models with Endogenous Regressors. In Richard W. Blundell, Whitney K. Newey, and Torsten Persson, editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume 3, pages 17–46. Cambridge University Press, Cambridge, United Kingdom, 2007. 5

Guido W. Imbens and Joshua D. Angrist. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475, 1994. 2, 7

Guido W. Imbens and Whitney K. Newey. Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, 77(5):1481–1512, 2009. 3, 5, 12

Sung Jae Jun. Local Structural Quantile Effects in a Model with a Nonseparable Control Variable. *Journal of Econometrics*, 151(1):82–97, 2009. 8, 12

Charles F Manski. Anatomy of the Selection Problem. *The Journal of Human Resources*, 24(3):343–360, 1989. 7

Matthew A. Masten. Random Coefficients on Endogenous Variables in Simultaneous Equations Models. *Working paper*, 2012. 10

Matthew A. Masten and Alexander Torgovitsky. Instrumental variables estimation of a generalized correlated random coefficients model. *cemmap working paper CWP02/14*, 2014. 5, 12

Christoph Rothe. Semiparametric Estimation of Binary Response Models with Endogenous Regressors. *Journal of Econometrics*, 153(1):51–64, November 2009. ISSN 0304-4076. URL http://www.sciencedirect.com/science/article/pii/S0304407609001110. 12

Alexander Torgovitsky. Minimum Distance from Independence Estimation of Nonseparable Instrumental Variables Models. *Working paper*, 2013. 12

Alexander Torgovitsky. Identification of nonseparable models using instruments with small support. *Econometrica*, 83(3):1185–1197, 2015. 5, 10

Jeffrey M Wooldridge. On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model. *Economics Letters*, 56(2):129–133, 1997. 3, 9

Jeffrey M. Wooldridge. Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model. *Economics Letters*, 79 (2):185–191, May 2003. 3, 9

Jeffrey M. Wooldridge. Instrumental Variables Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model. In Tom Fomby, R. Carter Hill, Daniel L. Millimet, Jeffrey A. Smith, and Edward J. Vytlacil, editors, *Modelling and Evaluating Treatment Effects in Econometrics (Advances in Econometrics, Volume 21)*, pages 93–116. Emerald Group Publishing Limited, 2008. 3, 9

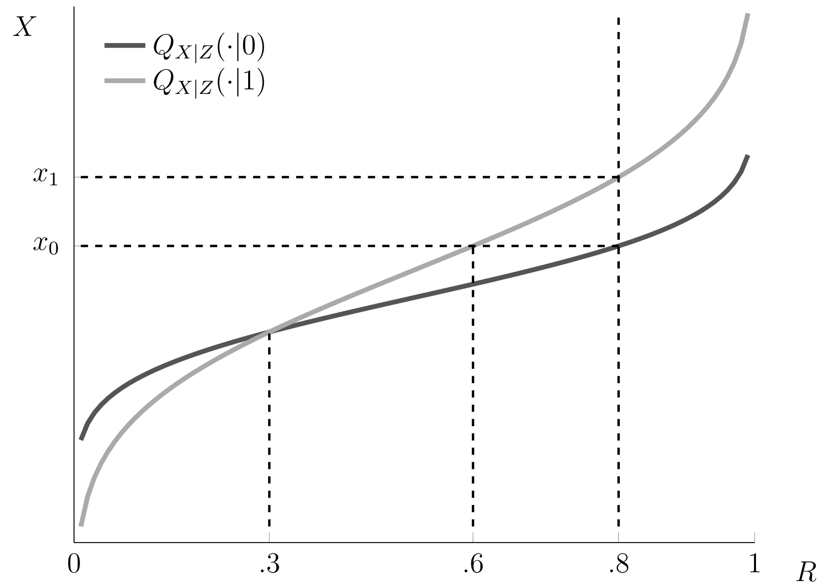

Figure 1: Consider the simple CRC model (1), and suppose that $Z \in \{0, 1\}$. The figure plots the quantile functions of X given Z, $Q_{X|Z}(\cdot|z)$, for $z = 0$ and 1. Conditional on $R = 0.8$, X assumes two values, $x_0 \equiv Q_{X|Z}(.8|0)$ and $x_1 \equiv Q_{X|Z}(.8|1)$, depending on the realization of Z. Since $Z \perp\!\!\!\perp B \mid \{R = 0.8\}$, a mean regression of Y on X conditional on $R = 0.8$ identifies the means of the intercept and slope coefficients, $\mathbb{E}[B_0 \mid R = 0.8]$ and $\mathbb{E}[B_1 \mid R = 0.8]$. In this plot, the relevance condition I5 holds for almost every $r \in (0, 1)$, since the conditional quantile functions intersect only at $r = 0.3$. Hence, $\mathbb{E}[B \mid R = r]$ is identified for all $r \in (0, 1)$. Averaging then identifies $\mathbb{E}[B]$. Note that Z can have a non-monotonic effect on X. In this figure, its effect is positive for units with large R (above $R = 0.3$) and negative for units with small R.