The International Journal of Biostatistics

Volume 5, *Issue* 1 | 2009 | *Article* 22

Semiparametrically Efficient Estimation of Conditional Instrumental Variables Parameters

Maximilian Kasy*

*University of California, Berkeley, maxkasy@econ.berkeley.edu

Semiparametrically Efficient Estimation of Conditional Instrumental Variables Parameters*

Maximilian Kasy

Abstract

In this paper, I propose a set of parameters designed to identify the slope of structural relationships based on a combination of conditioning on covariates and the use of an exogenous instrument. After giving structural interpretations to these parameters in the context of specific semiparametric models, I derive their efficient influence curves in a fully nonparametric context as well as under imposition of restrictions on the instrument. These influence curves give the semiparametric efficiency bounds for regular asymptotically linear estimators of the parameters and allow the construction of asymptotically efficient estimators. Monte Carlo experiments finally demonstrate the good finite sample performance of such estimators.

KEYWORDS: instrumental variables, efficient influence curve

*Recipient of a DOC fellowship of the Austrian Academy of Sciences at the Department of Economics, UC Berkeley. I'd like to thank Mark van der Laan, Bryan Graham and James Powell at UC Berkeley as well as an anonymous referee for valuable comments.

1 Introduction

Two broad classes of identification strategies for causal parameters, relating outcomes to "treatments", are commonly used in disciplines using observational data such as Labor Economics and Public Health. First, using variation of treatment orthogonal to a set of observable covariates, justified by variants of conditional independence assumptions. Second, using variation of treatment driven by some exogenous "instrument". Applied papers often combine these two approaches, estimating parametric regressions using controls as well as instruments for the treatment of interest. This is justified in particular if the instrument is credibly exogenous only conditional on covariates. [Angrist & Pischke(2009)] p175f cite as an example the literature on the Vietnam war draft lottery studying the health and labor market impact on the men being sent to Vietnam - seminal concerning the use of instrumental variables for both Public Health and Labor Economics. In this literature, date of birth is a valid instrument for being sent to Vietnam only conditional on year of birth, since a sequence of dates of birth was drawn randomly according to which men were drafted but different cutoffs in this sequence were applied for different cohorts.

In this paper I provide a semiparametric generalization of the approach combining conditioning and instruments, allowing for conditioning covariates to enter nonparametrically while staying linear in spirit concerning the use of the instrument.

Throughout I will consider crossection data containing a treatment variable T, an outcome variable Y, an "exogenous" instrument I as well as covariates X. The parameters (defined here in a purely statistical - as opposed to structural or causal - way) that I will consider are:

Definition 1 *Conditional instrumental variable regression (CIV):*

$$\beta_{CIV} := E\left[\frac{Cov(Y, I|X)}{Cov(T, I|X)}\right]$$

Definition 2 *Unconditional instrumental variable regression with estimated instrument (EIV):*

$$\beta_{EIV} := \frac{Cov(E[T|I], Y)}{Cov(E[T|I], T)}$$

Definition 3 *Conditional instrumental variable regression with estimated instrument (CEIV):*

$$\beta_{CEIV} := E\left[\frac{Cov(E[T|I, X], Y|X)}{Cov(E[T|I, X], T|X)}\right]$$

CIV generalizes linear multivariate IV (instrumental variables) regression of Y on T and X instrumented with I and X. For EIV, the parametric analogon is two stage least squares regression of Y on T, instrumenting T with a number of nonlinear transformations of I. CEIV, finally, generalizes the combination of the two. In the next section I will provide structural interpretations of these parameters in the context of specific semiparametric models. These interpretations are similar in spirit to the ones developed in [Imbens & Angrist(1994)] and [Card(2001)].

The central contribution of this paper lies in the derivation of the efficient influence curves for these three parameters in the fully nonparametric (or "saturated") model in section 3 as well as under imposition of restrictions on the distribution of I given X in section 4. These influence curves imply the semiparametric efficiency bounds for any regular asymptotically linear estimator of the respective parameters and can be used for the construction of estimators asymptotically achieving these bounds. For a general overview of the theory of semiparametric efficiency, compare [Tsiatis(2006)] or [Van der Laan & Robins(2003)].

In section 5, I outline the construction of efficient estimators. One possibility is based on the estimating function approach, using the efficient influence curve (with nonparametrically estimated nuisance parameters) as moment criterion function. Just as the use of the score of parametric models as criterion function yields estimators (maximum likelihood!) asymptotically achieving the Fischer efficiency bound, the use of the efficient score in semi- or nonparametric (infinite dimensional) models as criterion yields estimators asymptotically achieving the semiparametric efficiency bound.

An alternative to the use of estimating functions is given by the recently developed targeted maximum likelihood (TMLE) approach, described for instance in [Van der Laan & Rubin(2006)]. This approach estimates a fully specified model of the joint distribution of all observables, optimally trading off variance and bias of the parameter of interest implied by this model using the efficient influence curve. Section 5.2 elaborates.

It should be emphasized that either of these estimation approaches requires, explicitly or implicitly, nonparametric estimation of several conditional expectation functions given covariates. This constitutes a lesser problem if the covariates are a low dimensional vector, such as for instance the one dimensional year of birth variable in the case of the Vietnam war draft lottery. If the vector of covariates is high dimensional, however, attempts at nonparametric estimation are likely to break down (due to the "curse of dimensionality") and the proposed estimators might be infeasible. As a referee pointed out to me, [Robins & Ritov(1997)] develop some general asymptotic

DOI: 10.2202/1557-4679.1153

theory for that case.

Concluding the paper, section 6 provides some selected Monte Carlo evidence illustrating the good finite sample performance of some of the proposed estimators.

2 Structural interpretations of the parameters

Let me now provide some specific structural models in which the above parameters have a useful interpretation. These models are generalizations of random coefficient models commonly used in the literature. An excellent overview of such models can be found in [Card(2001)]. Assume, first, that the following assumptions hold (where β and π are random variables as well):

$$Y = \beta_0 + \beta_1 T \tag{1}$$

$$T = \pi_0 + \pi_1 I \tag{2}$$

$$I \perp (\beta, \pi)|X \tag{3}$$

This is a "triangular system" model with additive and multiplicative (slope) heterogeneity in the structural functions that imposes linearity of the structural functions. Under these assumptions we have

$$Cov(Y, I|X) = Cov(\beta_0 + \beta_1(\pi_0 + \pi_1 I), I|X) =$$

$$= Cov(\beta_1 \pi_1 I, I|X) = E[\beta_1 \pi_1 |X] Var(I|X) \tag{4}$$

where the last two equalities follow from the conditional exogeneity of I given X relative to the parameters π and β. Similarly

$$Cov(T, I|X) = E[\pi_1|X] Var(I|X) \tag{5}$$

Defining the weight function

$$\tilde{\pi} := \frac{\pi_1}{E[\pi_1|X]}$$

this implies that β_{CIV} recovers a weighted average of slope parameter β_1, where the weights integrate to one conditional on X:

$$\beta_{CIV} = E\left[\frac{Cov(Y, I|X)}{Cov(T, I|X)}\right] = E[\beta_1 \tilde{\pi}] \tag{6}$$

Now, let us turn to another model that relaxes the linearity of the first stage. For expositional simplicity I restrict to additive heterogeneity, multiplicative heterogeneity would be equally admissible. Assume in particular

$$\begin{aligned} Y &= \beta_0 + \beta_1 T && (7) \\ T &= g(I) + \epsilon && (8) \\ I &\perp (\beta, \epsilon)|X && (9) \end{aligned}$$

Then we have, similar to before,

$$\begin{aligned} &Cov(Y, E[T|I,X]|X) = Cov(\beta_0 + \beta_1(g(I)+\epsilon), g(I) + E[\epsilon|X]|X) = \\ &= Cov(\beta_1 g(I), g(I)|X) = E[\beta_1|X]Var(g(I)|X) \qquad (10) \end{aligned}$$

and

$$Cov(T, E[T|I,X]|X) = Var(g(I)|X) \qquad (11)$$

implying

$$\beta_{CEIV} = E\left[\frac{Cov(E[T|I,X], Y|X)}{Cov(E[T|I,X], T|X)}\right] = E[\beta]$$

that is, we recover the population average slope of the structural function relating Y to T.

If we generalize the model to allow for multiplicative heterogeneity in the structural relation between T and I, we recover weighted averages similar to the linear model discussed above.

3 Influence curves in the nonparametric setup

In this section and the following two the central results are presented, namely the influence curves of the various parameters defined above under the different restrictions on the data generating process. In particular, I will first consider the fully nonparametric model assuming (X, I, T, Y) have a smooth joint density which is otherwise left unrestricted. I then assume that the instrument is fully exogenous, i.e. $I \perp X$ is imposed. Finally, it is assumed that $p(I|X)$ is known. All the proofs are relegated to appendix A.

3.1 The efficient influence function for conditional IV with given instrument

We can factor the likelihood of the observed data as

$$p(x,i,t,y) = p(x)p(i,t,y|x)$$

and hence $\beta_{CIV} =$

$$\int p(x) \left[\frac{\int iyp(i,t,y|x)didtdy - \int ip(i,t,y|x)didtdy \int yp(i,t,y|x)didtdy}{\int itp(i,t,y|x)didtdy - \int ip(i,t,y|x)didtdy \int tp(i,t,y|x)didtdy} \right] dx \tag{12}$$

In order to proceed in deriving the semiparametric efficiency bound for the asymptotic variance of regular asymptotically linear estimators of β_{CIV}, we have to find its efficient influence function which - since we have a fully nonparametric (saturated) model - is the unique mean zero function ϕ satisfying

$$\frac{\partial \beta_{CIV}(\epsilon)}{\partial \epsilon} = E[\phi s]$$

for all possible score functions s of (x, i, t, y) (i.e. s square integrable with mean zero), where $\beta_{CIV}(\epsilon)$ is understood to be the parameter for the model $p(1 + \epsilon s)$. For an introduction to semiparametric theory and proof of this assertion compare [Tsiatis(2006)] or [Van der Laan & Robins(2003)].
Define now

$$\bar{\beta}(x) = \frac{Cov(Y, I|X)}{Cov(T, I|X)}$$

Then, based on this notion of the efficient influence function, we have the following

Theorem 1 *The unique influence function of any RAL estimator of β_{CIV} in the fully nonparametric model is given by*

$$\phi_{CIV}(X, I, T, Y) =$$

$$\bar{\beta}(x) \left(1 + \left[\frac{(Y - E[Y|X])(I - E[I|X])}{Cov(Y, I|X)} \right] - \left[\frac{(T - E[T|X])(I - E[I|X])}{Cov(T, I|X)} \right] \right) - \beta_{CIV}$$

3.2 The efficient influence function for unconditional IV using the estimated instrument $E[T|I]$

The result of the previous section obviously translates directly to the case where we replace I with $k(I)$ for arbitrary, known k. But what if we have to estimate k, in particular when we choose the instrument $k(I) := E[T|I]$? This would correspond to the notion of the optimal instrument in the case of a linear second stage with homoscedastic errors that are mean independent of the instrument I.

Theorem 2 *The unique influence function of any RAL estimator of* β_{EIV} *in the fully nonparametric model is given by*

$$\phi_{EIV}(I,T,Y) =$$
$$\beta_{EIV}\left(\left[\frac{(Y-E[Y])(E[T|I]-E[T])}{Cov(Y,E[T|I])}\right]-\left[\frac{(T-E[T])(E[T|I]-E[T])}{Cov(T,E[T|I])}\right]\right.$$
$$\left.+\left[\frac{(E[Y|I]-E[Y])(T-E[T|I])}{Cov(Y,E[T|I])}\right]-\left[\frac{(E[T|I]-E[T])(T-E[T|I])}{Cov(T,E[T|I])}\right]\right)$$

3.3 The efficient influence function for conditional IV using the estimated instrument $E[T|I,X]$

Finally, we can combine conditioning and estimation of the instrument to obtain β_{CEIV}

Let us furthermore denote

$$\bar{\beta}_{CEIV}(x) := \frac{Cov(E[T|I,X=x],Y|X=x)}{Cov(E[T|I,X=x],T|X=x)}$$

Then we have

Theorem 3 *The unique influence function of any RAL estimator of* β_{CEIV} *in the fully nonparametric model is given by*

$$\phi_{CEIV}(I,T,Y,X) = -\beta_{CEIV} + \bar{\beta}_{CEIV}(X)\times$$
$$\left(1+\left[\frac{(Y-E[Y|X])(E[T|I,X]-E[T|X])}{Cov(Y,E[T|I,X]|X)}\right]-\left[\frac{(T-E[T|X])(E[T|I,X]-E[T|X])}{Cov(T,E[T|I,X]|X)}\right]\right.$$
$$\left.+\left[\frac{(E[Y|I,X]-E[Y|X])(T-E[T|I,X])}{Cov(Y,E[T|I,X]|X)}\right]-\left[\frac{(E[T|I,X]-E[T|X])(T-E[T|I,X])}{Cov(T,E[T|I,X]|X)}\right]\right)$$

4 Influence curves imposing restrictions on the instrument

4.1 Imposing independence of I and X

As before, any probability density can be decomposed into

$$p(x,i,t,y) = p(x)p(i|x)p(t,y|i,x).$$

Imposing $I \perp X$ means imposing $p(i|x)$ to be constant in x. The scores of the restricted model can then be decomposed into $s(x,i,t,y) = s_1(x) + s_2(i) + s_3(t,y|x,i)$.

DOI: 10.2202/1557-4679.1153

Lemma 1 *The tangent space of the model imposing independence of I and X is spanned by these scores and is given by the orthocomplement of S_o where*

$$S_o = \{s_o(X, I) : E[s_o|X] = 0 \text{ and } E[s_o|I] = 0\}$$

Lemma 2 *The projection of any mean zero L^2 random variable s on S_o is given by*

$$s_p = \Pi_{S_o} s = E[s|X, I] - E[s|X] - E[s|I]$$

The projection of any influence function on the tangent space is now given by this influence function minus its projection on S_o. From semiparametric theory we know that to find the influence functions for the restricted model we have to find the projections of the nonparametric-model influence function on the tangent space. This is what underlies the following result.

Theorem 4 *Under the restriction $I \perp X$, the efficient influence curve for β_{CIV} is given by*

$$\psi_{CIV} = \bar{\beta}(X) - \beta_{CIV} + \chi + E[\chi|I]$$

where

$$\chi = \bar{\beta}(X) \left(\left[\frac{(Y - E[Y|X, I])(I - E[I|X])}{Cov(Y, I|X)} \right] - \left[\frac{(T - E[T|X, I])(I - E[I|X])}{Cov(T, I|X)} \right] \right)$$

and for β_{CEIV} by

$$\psi_{CEIV} = \bar{\beta}_{CEIV}(X) - \beta_{CEIV} + \rho + E[\rho|I] + \sigma$$

where

$$\rho = \bar{\beta}_{CEIV}(X) \left(\left[\frac{(Y - E[Y|X, I])(E[T|I, X] - E[T|X])}{Cov(Y, E[T|I, X]|X)} \right] - \left[\frac{(T - E[T|X, I])(E[T|I, X] - E[T|X])}{Cov(T, I|X)} \right] \right)$$

and

$$\sigma = \left[\frac{(E[Y|I, X] - E[Y|X])(T - E[T|I, X])}{Cov(Y, E[T|I, X]|X)} \right] - \left[\frac{(E[T|I, X] - E[T|X])(T - E[T|I, X])}{Cov(T, E[T|I, X]|X)} \right]$$

4.2 Assuming $p(I|X)$ known

If $p(i|x)$ is known without any further model restrictions imposed, then the orthocomplement of the tangent space is given by

$$S_o = \{s(X, I) : E[s|X] = 0\}$$

As can easily be verified, projection on S_o is given by

$$s_p = \Pi_{S_o} s = E[s|X, I] - E[s|X]$$

By the same argument as before we can then calculate the efficient influence functions in this model.

Theorem 5 *Under the restriction that $p(I|X)$ is given, the efficient influence curve for β_{CIV} is given by*

$$\tau_{CIV} = \bar{\beta}(x)\left(1 + \left[\frac{(Y - E[Y|X, I])(I - E[I|X])}{Cov(Y, I|X)}\right] - \left[\frac{(T - E[T|X, I])(I - E[I|X])}{Cov(T, I|X)}\right]\right) - \beta_{CIV}$$

and for β_{CEIV} by

$$\tau_{CEIV} = -\beta_{CEIV} + \bar{\beta}(x)\cdot$$
$$\cdot\left(1 + \left[\frac{(Y - E[Y|X, I])(E[T|I, X] - E[T|X])}{Cov(Y, E[T|I, X]|X)}\right] - \left[\frac{(T - E[T|X, I])(E[T|I, X] - E[T|X])}{Cov(T, E[T|I, X]|X)}\right]\right.$$
$$\left. + \left[\frac{(E[Y|I, X] - E[Y|X])(T - E[T|I, X])}{Cov(Y, E[T|I, X]|X)}\right] - \left[\frac{(E[T|I, X] - E[T|X])(T - E[T|I, X])}{Cov(T, E[T|I, X]|X)}\right]\right)$$

5 Estimation

5.1 Using estimating functions

A general strategy for constructing asymptotically efficient estimators given that we know the efficient influence curve is based on simply turning the condition $E[\phi] = 0$ into the estimating equation $E_n[\hat{\phi}(\hat{\beta})] = 0$, where ϕ the respective influence function and $\hat{\beta}$ is the respective parameter of interest. $\hat{\phi}$ in general contains terms corresponding to unknown parameters of the underlying model, which have to be estimated. It can easily be seen that this yields estimators which are regular, asymptotically linear and efficient. Proofs can be found for instance in [Van der Laan & Robins(2003)] or [Tsiatis(2006)].

This approach is possible only insofar as the parameter of interest appears in the influence function. In this paper this is the case for the conditional IV estimators with or without estimated instrument in the unrestricted model and the model imposing independence of X and I or known $p(I|X)$.

The estimating function approach is not applicable in the unconditional cases, as the parameter of interest does not appear in the influence function.

Construction of the estimating function in our cases requires estimation of a number of conditional moments. In the case of CIV those are $E[I|X]$, $E[T|X]$, $E[Y|X]$, $Cov(I,T|X)$ and $Cov(I,Y|X)$. I propose to estimate first the three conditional expectations and then, using these first stage estimates, to estimate the conditional covariances. A host of different nonparametric estimators could be used, in particular kernel and sieve estimators. Consistency of the estimators of all of these conditional moments is required for asymptotically efficient estimation. However, as the following result implies - by the general argument given for instance in [Tsiatis(2006)], less is required for root n consistent, asymptotically normal estimation - in other words the estimating function is robust to certain misspecifications, in particular either $E[I|X]$ and $Cov(Y,I|X)$ or $E[T|X]$, $E[Y|X]$ and $Cov(Y,I|X)$ can be inconsistently estimated without affecting convergence of the estimator of interest.

Proposition 1

$$E[\hat{\phi}_{CIV}(\beta_{CIV})] = 0$$

if either

$$\hat{E}[I|X] = E[I|X] \text{ and } \hat{Cov}(T,I|X) = Cov(T,I|X) \tag{13}$$

or

$$\hat{E}[T|X] = E[T|X],\ \hat{E}[Y|X] = E[Y|X] \text{ and } \hat{Cov}(T,I|X) = Cov(T,I|X) \tag{14}$$

5.2 Targeted maximum likelihood

We can now furthermore proceed to define a targeted maximum likelihood estimator along the lines proposed recently in [Van der Laan & Rubin(2006)]. For a formal, self contained discussion, the reader is referred to the cited paper. The general idea is to estimate a fully specified, generally parametric but flexible, model of the joint density of all variables. In contrast to ordinary maximum likelihood estimation, however, a somewhat different objective function is maximized that "targets" the parameter of interest, optimally trading of bias and variance of this parameter. As [Van der Laan & Rubin(2006)] show, the following procedure achieves the semiparametric efficiency bound (assuming correct specification of nuisance parameters) and delivers an estimate of a fully specified model.

1. Estimate an initial (parametric) model $\hat{p}_1$ specifying the full joint likelihood of (X, I, T, Y)

2. Construct a family of models parametrized by ϵ of the form $\hat{p}_1(1 + \epsilon\phi)$, where the nuisance parameters in ϕ are calculated from $\hat{p}_1$. ϕ is the respective efficient influence curve derived above, where the choice of ϕ depends on the parameter of interest and the assumptions imposed on I.

3. Calculate the MLE of this family to obtain an updated model

4. Iterate steps 2 and 3 with the updated density.

Asymptotically, efficiency is achieved with only one iteration, in finite samples several iterations will improve performance.

In any parametric model that we might choose, we want to be careful for allowing the nuisance parameters to enter flexibly in order to get the efficient influence curves right. For CIV, a set of models that achieves this might be as follows:

- Take some general purpose model for the unconditional distribution of X, as fit in the given context, possibly use a nonparametric kernel density estimator
- Assume $(I, T, Y)|X \sim N(\mu(X), \Sigma(X))$
- Specify $\mu(X)$ and $\Sigma(X)$ by some flexible series, for instance as polynomials.

If the specification of μ and Σ is flexible enough, all nuisance parameters that enter the influence curve for β_{CIV} (namely $E[(I, T, Y)|X]$, $Cov(T, I|X)$ and $Cov(Y, I|X)$) should be consistently estimated even if the normality assumption is violated, since the MLE for conditional normals gets conditional expectations and variances right under misspecification, assuming the imposed models for μ and Σ asymptotically contain the truth. As a consequence the corresponding TMLE of β_{CIV} will be consistent and achieve the semiparametric efficiency bound.

Let's now look at β_{EIV}. Choosing a model flexible enough to get all the relevant (nuisance) parameters right is somewhat more tricky here. For efficient estimation we need to consistently estimate $E[T|I]$, $E[Y|I]$, $Cov(Y, E[T|I])$ as well as $Cov(T, E[T|I])$. The following type of model achieves this:

- Assume $(T, Y)|I \sim N(\mu(I), \Sigma(I))$, with, in particular, $\mu(I)$ polynomial of order k

DOI: 10.2202/1557-4679.1153

- Choose a family of densities for I that allows the first $2k$ moments to be estimated consistently, such as a nonparametric kernel density estimator.

Consistency of the first $2k$ moments is required because of the covariances, which in the case of polynomial of order k conditional means are a (linear) function of these $2k$ moments.

β_{CEIV} requires similar considerations as β_{EIV}, conditional on X. Furthermore, we need again some general purpose model for X. The following should deliver consistent estimators of the relevant nuisance parameters:

- Assume $(T,Y)|I,X \sim N(\mu(I,X),\Sigma(I,X))$, with, in particular, $\mu(I,X)$ polynomial of order k in both X and I
- Choose a family of densities for (X,I) that allows the first $2k$ moments of I conditional on X to be consistently estimated. A kernel density estimator seems again a reasonable choice.

Finally, note that we can use the same models if we impose a priori that $I \perp X$ or $p(I|X)$ is known. While direct imposition of these restrictions will improve efficiency of the first stage, and hence seems quite reasonable in practice, asymptotic performance of the TMLE for our various parameters of interest should be unaffected by whether or not the restrictions are imposed, since asymptotic efficiency only requires consistency of the first stage.

6 Some Monte Carlo evidence

To illustrate the practical performance of estimators based upon the influence curves derived, I will present Monte Carlo evidence in this section, without however attempting a systematic evaluation of the finite sample properties of all estimators proposed which would be beyond the scope of the present paper. I will restrict discussion to estimation of β_{CIV} using the estimating function approach. The Matlab/Octave code for the estimators and for the simulations are available from the author upon request.

For all simulation designs, I estimate the parameter β_{CIV} using the efficient influence curve given knowledge of $P(I|X)$, using the nonparametric efficient influence curve, using the "naive" estimator $E[\hat{\beta}(X)]$ and using the unconditional two stage least squares estimator $\hat{Cov}(I,Y)/\hat{Cov}(I,T)$. The designs use uniformly or normally distributed X, with I,T and Y being jointly normally distributed conditional on X. The conditional means and covariances of I,T and Y are chosen to be polynomials in X or smooth functions of such polynomials.

As it turned out in the simulations, the fact that estimated conditional covariances appear in the denominators of the various terms averaged to obtain the respective estimators leads to ill behaved estimators due to "outliers" generated by almost-zero estimated denominators. This problem bears some resemblance to estimators using inverse probability weighting, where estimated densities appear in the denominator. The solution to this problem proposed in the literature is trimming, i.e. averaging over the subset of the support of the conditioning variables on which the denominator is bounded away from zero by some constant. If the true propensity scores are not bounded away from zero, this affects rates of convergence and asymptotic distribution theory more generally.

I use a related approach, based on symmetrically "censoring" the estimating functions around an initial estimate. The estimators are hence censored means of $\hat{\beta}(X)$ for the naive estimator, censored means of $\phi + \hat{\beta}$ for the influence curve based estimators. Values of ϕ that are larger (or smaller) than a prespecified amount (30 times $\pm\hat{\beta}$ in my case) are replaced by that amount before calculating the mean. If the true conditional covariances are bounded away from zero, the asymptotic distribution should not be affected as long as asymptotically no terms are censored. A formal proof of this assertion is left for future research.

The following tables 1, 2 and 3 show means and variances for the various estimators under the designs used as well as the implied mean squared errors. Full model specifications for each design can be found in appendix B. Additional tables showing median and mean absolute deviation for these simulations can be found in appendix C. For ease of reading, the true β_{CIV} has been normalized to 1 in all designs.

Studying these tables reveals - exceptions notwithstanding - the following general patterns:

- The bias of the unconditional IV estimator is very large. This is not surprising given the choice of designs which illustrate the motivating point of this paper that exogeneity assumptions might hold conditionally but not unconditionally. Somewhat more surprisingly, the variance of the unconditional IV estimator is also much larger than the one of the conditional estimators, although this might be driven by small denominators in the designs used.

- For designs 1 and 2, conditioning on one dimensional X, the bias of the influence curve based estimator using information on $P(I|X)$ tends to be smaller then the bias of the fully nonparametric efficient estimator, which in turn has a smaller bias then the naive estimator. For designs 3

DOI: 10.2202/1557-4679.1153

and 4, conditioning on two dimensional X, the nonparametric efficient estimator tends to perform best in terms of bias.

- For the smaller of the sample sizes chosen, there does not appear a consistent pattern in terms of relative variances. For the larger sample sizes, variances become roughly similar, with the nonparametric estimator often performing best.

- The combination of these patterns implies that the mean squared error for the nonparametric efficient estimator tends to be smallest of the four estimators shown for most of the designs and sample sizes of 1000 and larger, although the naive estimator shows a fairly similar performance.

- Finally, the tables in the appendix displaying statistics that are more robust to outliers show a similar pattern. The nonparametric estimator consistently performs best in terms of median bias and is on a par with the naive estimator in terms of mean absolute deviation.

I take these results to suggest that the efficient influence curve based estimator not imposing any restrictions indeed tends to perform well as implied by asymptotic theory - at least given that the dimensionality of conditioning covariates is low and that the conditional means and covariances are smooth functions of the covariates. The estimator using additional knowledge of $P(I|X)$, on the other hand, appears to be hurt by the necessity of more demanding first stage nonparametric regressions.

Table 1: Bias of Estimators in Monte Carlo Simulations

design	n	using $P(I\|X)$	nonparametric	naive	unconditional
1	300	-0.5798	0.1794	0.1124	1.0027
	1,000	-0.0276	0.0409	0.0467	0.8837
	3,000	-0.0005	0.0128	0.0131	0.8613
	10,000	0.0010	0.0040	0.0041	0.8508
2	300	-0.1855	0.1165	0.1463	8.7849
	1,000	-0.0379	0.0297	0.0394	8.5778
	3,000	-0.0056	0.0116	0.0136	7.7459
	10,000	-0.0006	0.0034	0.0036	7.5538
3	300	-1.0746	-0.0281	0.1936	0.5384
	1,000	-0.2126	0.0664	0.1018	0.4677
	3,000	-0.0362	0.0267	0.0314	0.4499
	10,000	-0.0110	0.0072	0.0077	0.4437
4	300	-0.4108	0.0312	0.1402	5.1441
	1,000	-0.1223	0.0421	0.0701	4.2165
	3,000	-0.1818	-0.0141	0.0557	3.1938
	10,000	-0.0044	0.0077	0.0082	3.9397

Table 2: Variance of Estimators in Monte Carlo Simulations

design	n	using $P(I\|X)$	nonparametric	naive	unconditional
1	300	748.9410	57.8526	5.7851	0.8404
	1,000	0.0767	0.0334	0.0352	0.0906
	3,000	0.0102	0.0091	0.0091	0.0279
	10,000	0.0030	0.0025	0.0025	0.0080
2	300	1.6365	13.0942	17.3428	188.1762
	1,000	0.0542	0.0345	0.0320	22.0399
	3,000	0.0138	0.0093	0.0091	2.9138
	10,000	0.0038	0.0024	0.0024	0.7373
3	300	26.2986	11.0989	3.7751	0.3043
	1,000	0.3880	0.0746	0.0589	0.0539
	3,000	0.0122	0.0099	0.0102	0.0161
	10,000	0.0029	0.0025	0.0025	0.0047
4	300	9.0956	9.0313	5.3545	56.7395
	1,000	0.1955	0.1750	0.0822	2.1868
	3,000	0.1100	0.0283	0.0203	0.3004
	10,000	0.0031	0.0025	0.0025	0.1314

DOI: 10.2202/1557-4679.1153

Table 3: Mean Squared Error of Estimators in Monte Carlo Simulations

design	n	using $P(I\|X)$	nonparametric	naive	unconditional
1	300	350.1285	27.0522	2.7146	1.3978
	1,000	0.0366	0.0173	0.0186	0.8232
	3,000	0.0048	0.0044	0.0044	0.7549
	10,000	0.0014	0.0012	0.0012	0.7275
2	300	1.5269	11.9558	15.8385	248.7957
	1,000	0.0509	0.0324	0.0307	93.6791
	3,000	0.0126	0.0086	0.0085	62.6566
	10,000	0.0035	0.0022	0.0022	57.7329
3	300	12.2212	4.6712	1.6260	0.4179
	1,000	0.2084	0.0358	0.0351	0.2414
	3,000	0.0064	0.0049	0.0053	0.2092
	10,000	0.0013	0.0011	0.0011	0.1988
4	300	8.4735	8.2470	4.9086	78.2678
	1,000	0.1934	0.1616	0.0800	19.7756
	3,000	0.0927	0.0156	0.0141	10.3632
	10,000	0.0029	0.0024	0.0024	15.6409

7 Conclusion

In this paper I have introduced a series of conditional IV parameters that combine the intuitions from identification strategies for structural parameters using variation in treatment orthogonal to covariates (conditioning) and identification strategies using variation in treatment driven by covariates (instrumental variables). In the context of several structural models of the random coefficients type I give interpretations to these parameters as weighted conditional local average treatment effects, if "exogeneity" and "exclusion" of the instrument are satisfied conditional on covariates.

The central part of the paper then proceeds deriving the efficient influence curves for these parameters, both in a fully unrestricted (saturated) model and under restrictions on the distribution of the instrument conditional on covariates. These influence curves give, first, the efficiency bound for any regular asymptotically linear estimator. They allow, second, the construction of estimators that asymptotically achieve this efficiency bound. I sketch two approaches from the literature and show how they can be applied in the present context, based either on estimating functions or on the recently introduced targeted maximum likelihood approach. Both approaches require first stage nonparametric estimation of nuisance parameters and only yield useful estimators if the dimensionality of covariates is low, due to the "curse of dimensionality".

Finally, I conduct a series of Monte Carlo experiments that illustrate the performance of some of the estimators introduced. The estimators have to be slightly modified to account for outliers driven by small estimated denominators.

From these experiments I conclude that the efficient estimator based on the fully nonparametric influence curve has good finite sample properties within the range of models used in the simulations.

A Appendix - proofs

Proof of Theorem 1: Any score can be written as

$$s(x,i,t,y) = s(x) + s(i,t,y|x)$$

Let us first consider submodels of the form $p_\epsilon = p(1+\epsilon s(x))$ indexed by the one dimensional parameter ϵ. Subscripting by ϵ will henceforth indicate that the corresponding expressions are evaluated with respect to the probability distribution p_ϵ. Then

$$\begin{aligned}\frac{\partial \beta_{CIV}(\epsilon)}{\partial \epsilon} &= \int p(x)s(x)\left[\frac{Cov(Y,I|X)}{Cov(T,I|X)}\right]dx \\ &= E[(\bar{\beta}(x) - \beta_{CIV})s(x)] \qquad (15)\end{aligned}$$

where we define

$$\bar{\beta}(x) = \frac{Cov(Y,I|X)}{Cov(T,I|X)}$$

and $-\beta_{CIV}$ is added to the influence curve to ensure mean zero.

Next let us turn to the models $p_\epsilon = p(1+\epsilon s(i,t,y|x))$. For such models we have

$$\begin{aligned}\frac{\partial Cov_\epsilon(Y,I|X)}{\partial \epsilon} &= \frac{\partial}{\partial \epsilon}\left[\int iyp_\epsilon(i,t,y|x)didtdy\right. \\ &- \left.\int ip_\epsilon(i,t,y|x)didtdy \int yp_\epsilon(i,t,y|x)didtdy\right] \\ &= E[IYs|X] - E[Is|X]E[Y|X] - E[Ys|x]E[I|X] \\ &= E[(I - E[I|X])(Y - E[Y|X])s|X] \qquad (16)\end{aligned}$$

DOI: 10.2202/1557-4679.1153

where I took the liberty in the last line to insert the term $E[Y|X]E[I|X]s$ which has to have zero expectation conditional on X since s is a conditional score with conditional zero expectation.

Equation 16 then implies

$$\frac{\partial \bar{\beta}_\epsilon(x)}{\partial \epsilon} = \bar{\beta}(x) \left(\frac{\frac{\partial}{\partial \epsilon} Cov_\epsilon(Y, I|X)}{Cov(Y, I|X)} - \frac{\frac{\partial}{\partial \epsilon} Cov_\epsilon(T, I|X)}{Cov(T, I|X)} \right)$$

$$= E\left[\bar{\beta}(x) \left(\frac{(I - E[I|X])(Y - E[Y|X])}{Cov(Y, I|X)} - \frac{(I - E[I|X])(T - E[T|X])}{Cov(T, I|X)} \right) s|X\right] \tag{17}$$

whereby we have found the influence curve for variations in the conditional probabilities. As it also turns out, we have already arranged things such that the influence curve implicit in 17 has conditional mean zero. But now, by linear combination, we have covered all possible scores, hence combining equations 15 and 17 proofs the theorem. □

Proof of Theorem 2: As before, we consider the directional derivatives in direction s by looking at one dimensional submodels of the form $p(1 + \epsilon s)$. Let's first consider the numerator of β_{EIV} and remember that we set $k(I) := E[T|I]$:

$$\frac{\partial Cov_\epsilon(Y, E[T|I])}{\partial \epsilon} = \frac{\partial}{\partial \epsilon} \left[\int y \left[\int t p_\epsilon(t|i) \right] p_\epsilon(i, y) di dy - \int t p_\epsilon(t) dt \int y p_\epsilon(y) dy \right]$$

$$\begin{aligned}
&= E[k(I)Ys] - E[Ts]E[Y] - E[Ys]E[T] + E[YE[T(s - E[s|I])]] && (18)\\
&= E[(Y - E[Y])(k(I) - E[k(I))s] \\
&\quad + E[Y]E[(k(I) - T)s] + E[YE[T(s - E[s|I])]] \\
&= E[(Y - E[Y])(k(I) - E[k(I))s] \\
&\quad + E[(E[Y|I] - E[Y])(T - E[T|I])s] && (19)
\end{aligned}$$

where the last term in 18 reflects the fact that $k(.)$ has to be estimated and $s - E[s|I]$ is the score of $p_\epsilon(t|i)$. 19 then follows by noting that

$$-E[(k(I) - T)s] = E[Cov(T, s|I)] = E[T(s - E[s|I])]$$

By a similar argument

$$\begin{aligned}
\frac{\partial Cov_\epsilon(T, E[T|I])}{\partial \epsilon} &= E[(T - E[T])(k(I) - E[k(I))s] \\
&\quad + E[(E[T|I] - E[T])(T - E[T|I])s] && (20)
\end{aligned}$$

Now from 19 and 20 the claim is immediate.$\square$

Proof of Theorem 3: The claim follows from a straightforward combination of the arguments in the proofs of theorem 1 and 2. $\square$

Proof of Lemma 1: This follows since every element of the orthocomplement has to be uncorrelated with any function of X (i.e. any s_1), hence we need $E[s|X] = 0$, similarly $E[s|I] = 0$ and s has to be a function of X and I only in order to be orthogonal to any score s_3. Furthermore these conditions are sufficient.$\square$

Proof of Lemma 2: s_p lies in S_o: $E[s_p|X] = E[s|I] - E|s|I] - E[E[s|I]|X] = E[s] = 0$ where the second equality follows from independence of X and Y.
Furthermore, for any $s_o \in S_o$

$$E[(s - s_p)s_o] = E[(s - E[s|X, I] + E[s|X] + E[s|I])s_o]$$

$$= E[s_o E[(s - E[s|X, I])|X, I]] + E[E[s|X]E[s_o|X]] + E[E[s|I]E[s_o|I]] = 0$$

$\square$

Proof of theorem 4: Let us first consider conditional IV. We have to compute the three conditional expectations referred to in Lemma 2. Two of those are straightforward. $E[\phi_{CIV}|X] = \bar{\beta}(X) - \beta_{CIV}$ and

$$E[\phi_{CIV}|X, I] = -\beta_{CIV} + \bar{\beta}(X)\cdot$$

$$\cdot\left(1 + \left[\frac{(E[Y|I, X] - E[Y|X])(I - E[I|X])}{Cov(Y, I|X)}\right] - \left[\frac{(E[T|I, X] - E[T|X])(I - E[I|X])}{Cov(T, I|X)}\right]\right)$$

Finally, reminding ourselves of the independence of X and I, we get an expression that unfortunately does not simplify much for

$$E[\phi_{CIV}|I] =$$

$$E\left[\bar{\beta}(X)\left(\left[\frac{(E[Y|I, X] - E[Y|X])(I - E[I|X])}{Cov(Y, I|X)}\right] - \right.\right.$$

$$\left.\left. - \left[\frac{(E[T|I, X] - E[T|X])(I - E[I|X])}{Cov(T, I|X)}\right]\right)\right| I\right]$$

DOI: 10.2202/1557-4679.1153

As for CEIV, the result follows by the same argument. The σ term for the influence curve for OICV remains unchanged relative to the unrestricted case as it is already orthogonal to S_o - its expectation given I, X equals zero.□

Proof of theorem 5: Straightforward from the characterization of the orthocomplement of the tangent space and the proof of theorem 4. □

Proof of Proposition 1: Let's consider case 13, the other case is analogous. If $\hat{E}[I|X] = E[I|X]$, then

$$E[(Y - \hat{E}[Y|X])(I - \hat{E}[I|X])|X] = Cov(Y, I|X)$$

and

$$E[(T - \hat{E}[T|X])(I - \hat{E}[I|X])|X] = Cov(T, I|X)$$

(as a side remark, this implies the possibility of consistent estimation of $Cov(T, I|X)$, if we have correct model for it). Now, under assumption 13, take

$$E[\hat{\phi}_{CIV}(\beta_{CIV})] =$$

$$= E\left[\frac{\hat{Cov}(Y,I|X)}{\hat{Cov}(T,I|X)}\left(1 + \left[\frac{(Y - \hat{E}[Y|X])(I - \hat{E}[I|X])}{\hat{Cov}(Y,I|X)}\right]\right.\right.$$

$$\left.\left. - \left[\frac{(T - \hat{E}[T|X])(I - \hat{E}[I|X])}{\hat{Cov}(T,I|X)}\right]\right) - \beta_{CIV}\right]$$

$$= E\left[\frac{\hat{Cov}(Y,I|X)}{Cov(T,I|X)}\left(1 + \left[\frac{Cov(Y,I|X)}{\hat{Cov}(Y,I|X)}\right] - \left[\frac{Cov(T,I|X)}{Cov(T,I|X)}\right]\right) - \beta_{CIV}\right]$$

$$= E\left[\frac{Cov(Y,I|X)}{Cov(T,I|X)} - \beta_{CIV}\right] = 0$$

□

B Appendix - Monte Carlo designs

The Monte Carlo designs used in section 6 are as follows:
For all designs

$$(I,T,Y)|X \sim N(\mu(X),\Sigma(X))$$

Furthermore, in all designs $\Sigma(X)_{I,T}$ and $\Sigma(X)_{I,Y}$ - which determine $\beta(X)$ - are specified first, the other components are chosen to give a positive definite covariance matrix:

$$\Sigma(X)_{I,I} = |\Sigma(X)_{I,T}| + |\Sigma(X)_{I,Y}|$$

$$\Sigma(X)_{T,T} = 16 \cdot \Sigma(X)^2_{I,T}/\Sigma(X)_{I,I}$$

$$\Sigma(X)_{Y,Y} = 16 \cdot \Sigma(X)^2_{I,Y}/\Sigma(X)_{I,I}$$

$$\Sigma(X)_{T,Y} = .25\sqrt{\Sigma(X)_{T,T}\Sigma(X)_{Y,Y}}$$

these choices imply conditional correlations of .25 between all three variables.

Design 1:

$$X \sim U([0,1])$$

$$\mu(X) = (1, X, X^2)\beta$$

with

$$\beta = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 0 \\ 1 & 3 & 3 \end{pmatrix}'$$

$$\Sigma(X)_{I,T} = (1, X, X^2) \cdot (5\ 0\ 1)'$$

$$\Sigma(X)_{I,Y} = (1, X, X^2) \cdot (3\ 0\ 2)'$$

Design 2:

$$X \sim N(0,1)$$

$\mu(X)$ as in design 1,

$$\Sigma(X)_{I,T} = atan((1, X, X^2) \cdot (5\ 0\ 1)')$$

$$\Sigma(X)_{I,Y} = atan((1, X, X^2) \cdot (3\ 0\ 2)')$$

DOI: 10.2202/1557-4679.1153

Design 3:

$$X \sim U([0,1]^2)$$

$$\mu(X) = XP \cdot \beta$$

where

$$XP = (1, X, X_1^2, X_2^2, X_1 X_2)$$

and

$$\beta = \begin{pmatrix} 3 & 2 & 4 & .1 & .3 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & .2 & .4 & .1 \end{pmatrix}'$$

$$\Sigma(X)_{I,T} = XP \cdot \beta_{IT}$$

$$\Sigma(X)_{I,Y} = XP \cdot \beta_{IY}$$

with

$$\beta_{IT} = (5\ 0\ 0\ 2\ 0\ 0)'$$

$$\beta_{IY} = (3\ 0\ 0\ 1\ 1\ 0)'$$

Design 4:

$$X \sim N((0,0), I_2)$$

$\mu(X)$ as in design 3,

$$\Sigma(X)_{I,T} = atan(XP \cdot \beta_{IT})$$

$$\Sigma(X)_{I,Y} = atan(XP \cdot \beta'_{IY})$$

C Additional statistics for Monte Carlos

Tables 4 and 5 show additional, robust statistics for the same set of simulations discussed in section 6.

Table 4: MEDIAN BIAS OF ESTIMATORS IN MONTE CARLO SIMULATIONS

design	n	using $P(I\|X)$	nonparametric	naive	unconditional
1	300	-0.1278	0.0215	0.0587	0.8374
	1,000	-0.0245	0.0230	0.0275	0.8466
	3,000	-0.0042	0.0074	0.0080	0.8486
	10,000	-0.0003	0.0029	0.0027	0.8469
2	300	-0.1296	0.0061	0.0356	6.9717
	1,000	-0.0377	0.0108	0.0213	7.4853
	3,000	-0.0086	0.0065	0.0083	7.4724
	10,000	-0.0015	0.0020	0.0021	7.4705
3	300	-0.5136	-0.0460	0.0726	0.4443
	1,000	-0.1388	0.0434	0.0746	0.4457
	3,000	-0.0351	0.0216	0.0253	0.4436
	10,000	-0.0121	0.0062	0.0065	0.4419
4	300	-0.3149	-0.0240	0.0562	3.8656
	1,000	-0.1027	0.0264	0.0528	3.9029
	3,000	-0.1453	-0.0119	0.0462	3.1280
	10,000	-0.0064	0.0065	0.0069	3.9167

Table 5: MEAN ABSOLUTE DEVIATION OF ESTIMATORS IN MONTE CARLO SIMULATIONS

design	n	using $P(I\|X)$	nonparametric	naive	unconditional
1	300	1.0336	0.5145	0.4108	0.5083
	1,000	0.1531	0.1409	0.1434	0.2324
	3,000	0.0803	0.0755	0.0754	0.1323
	10,000	0.0432	0.0396	0.0396	0.0707
2	300	0.4366	0.4105	0.4027	7.2313
	1,000	0.1710	0.1395	0.1374	2.8537
	3,000	0.0924	0.0751	0.0751	1.2849
	10,000	0.0493	0.0392	0.0393	0.6730
3	300	1.2235	0.6455	0.5020	0.3649
	1,000	0.2505	0.1715	0.1724	0.1812
	3,000	0.0839	0.0783	0.0796	0.1010
	10,000	0.0428	0.0400	0.0401	0.0546
4	300	0.6458	0.5435	0.4408	3.3559
	1,000	0.2013	0.1706	0.1579	1.0512
	3,000	0.1493	0.0881	0.0969	0.4256
	10,000	0.0446	0.0403	0.0404	0.2872

DOI: 10.2202/1557-4679.1153

References

[Angrist & Pischke(2009)] Angrist, J. & Pischke, J. (2009). *Mostly harmless econometrics: an empiricist's companion.* Princeton Univ Pr.

[Card(2001)] Card, D. (2001). Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica* , 1127–1160.

[Imbens & Angrist(1994)] Imbens, G. & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *ECONOMETRICA* **62**, 467–467.

[Robins & Ritov(1997)] Robins, J. & Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16**(3), 285–319.

[Tsiatis(2006)] Tsiatis, A. (2006). *Semiparametric Theory and Missing Data.* Springer.

[Van der Laan & Robins(2003)] Van der Laan, M. & Robins, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality.* Springer.

[Van der Laan & Rubin(2006)] Van der Laan, M. & Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics* **2**(1), 11.