Constrained Conditional Moment Restriction Models

V. Chernozhukov, W. Newey, A. Santos

EMES 2022

INTRODUCTION

Shape restrictions have a central role in economics as both testable implications of classical theory and sufficient conditions for informative counterfactual predictions.

An important example is demand analysis where Slutzky conditions provide shape restrictions.

There are many other applications including constraining probabilities to be nonnegative in structural modeling.

Another important application is nonparametric instrumental variables estimation (NPIV) where shape restrictions may mitigate the ill-posed inverse problem; Chetverikov and Wilhelm (2017).

Shape restrictions are often equivalent to inequality restrictions on parameters of interest and on certain unknown functions.

Inference with inequality restrictions is difficult.

Such restrictions lead to discontinuities in limiting distributions as the inequality restrictions become binding.

This makes inference challenging due to non-pivotal and potentially unreliable pointwise asymptotic approximations.

Limit discontinuities make it difficult to construct confidence intervals with uniform coverage.

Of course uniform coverage is importance to confidence statements that are approximately valid across a range of parameter values where constraints are binding or not.

We address these challenges by conservatively enforcing inequality restrictions constraints as they become close to binding.

This conservative enforcement gives uniformity because the constraints get imposed with rising probability as the parameters get close to points where constraints bind.

Tests and confidence intervals may be conservative but powerful in exploiting the large amount of information that inequality restrictions can provide.

The test statistic is the difference of inequality restricted and unrestricted GMM objective functions.

Critical values obtained by bootstrapping while conservatively enforcing inequalities.

Confidence intervals obtained by inverting test.

Straightforward computations for linear instrumental variables (IV) model with linear inequalities.

Give general results for partially identified, nonlinear models.

Apply to conditional moment restriction models that encompasses parametric (Hansen, 1982), nonparametric (Newey and Powell, 1989, 2003), and semiparametric (Ai and Chen, 2003) instrumental variable models.

Also applies to functions of parameter estimates in these models.

General approach allows for partial identification.

We conduct inference on the causal effect of childbearing on female labor force participation for the instrumental variables approach of Angrist and Evans (1998).

We find that monotonicity of the local average treatment effect (LATE) in education is not rejected by the data and neither is monotonicity and negativity.

Further, imposing these shape restrictions yields narrower confidence intervals for the LATE at different schooling levels.

Obtain similar results for the partially identified average treatment effect (ATE), though the data is less informative about the ATE because of the low proportion of compliers.

These methods have been used by Torgovitsky (2019) to construct informative confidence intervals for partially identified state dependence parameters in the presence of unobserved heterogeneity.

Also, Kline and Walters (2021) used these methods to test shape constraints implied by a model of callback probabilities for employment applications.

Freyberger and Horowitz (2015) gave inference methods for shape restricted partially identified discrete IV models but those methods do not provide uniform inference.

Our results are complementary to those of Bugni et al. (2017) and Kaido et al. (2019), who provide uniformly valid procedures for subvector inference with partial identification.

Our analysis is further related to Santos (2012), Tao (2014), and Chen et al. (2011) who study inference on functionals of potentially partially identified structural functions, but do not allow for shape constraints.

INFERENCE

We focus first on linear IV estimation with equality and inequality restrictions .

The model is

$$Y_i = W_i'\theta_0 + \varepsilon_i,\ E[Z_i\varepsilon_i] = 0,\ (i = 1, \ldots, n),$$

where Y_i is a left-hand side endogenous variable, W_i is a vector of right hand-side variables, and Z_i is a vector of instrumental variables.

Assume throughout that data $X_i = (Y_i, W_i, Z_i)$ are i.i.d..

The null hypothesis of interest is

$$H_0 : F\theta_0 = f,\ G\theta_0 \leq g.$$

Base tests on the two step GMM estimator with optimal weighting matrix as in Hansen (1982) and White (1982).

Let $\bar{\theta}$ be an initial IV estimator of θ_0 and $\hat{\Omega}$ be a corresponding estimator of $E[Z_i Z_i' \varepsilon_i^2]$, such as

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2, \hat{\varepsilon}_i = Y_i - W_i' \bar{\theta}.$$

Let $\hat{g}(\theta) = (1/n) \sum_{i=1}^{n} Z_i (Y_i - W_i' \theta)$ be sample moments of products of instrumental variables and residuals and

$$\hat{Q}(\theta) = \sqrt{n} [\hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta)]^{1/2}.$$

Square root of usual GMM objective is convenient for asymptotic theory.

Restricted and unrestricted GMM estimators are

$$\hat{\theta} = \arg \min_{F\theta = f,\, G\theta \leq g} \hat{Q}(\theta), \ \hat{\theta}_u = \arg \min_{\theta} \hat{Q}(\theta).$$

Test statistic is

$$T = \hat{Q}(\hat{\theta}) - \hat{Q}(\hat{\theta}_u).$$

Construct critical value using the bootstrap.

Let recentered moment vector be

$$\hat{g}_i = Z_i \hat{\varepsilon}_i - \frac{1}{n} \sum_{j=1}^{n} Z_j \hat{\varepsilon}_j, \ \hat{\varepsilon}_i = y_i - W_i' \hat{\theta}.$$

Let $b \in \{1, ..., B\}$ index bootstrap draws, $w_1^b, \ldots, w_n^b$ be i.i.d. $N(0, 1)$ independent of the data, and

$$W^b = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i^b \hat{g}_i$$

be a multiplier bootstrap draw of the moments.

To get critical value we impose a local version of the restrictions in the bootstrap.

h with the same dimension as θ serves as a possible value of $\sqrt{n}(\hat{\theta} - \theta_0)$.

$r_n > 0$ is a scalar slackness choice.

Local constrained parameter space is

$$\hat{V}_n = \left\{h : Fh = 0,\ G_j h \le \max\left\{0, -\sqrt{n}\left\{r_n + G_j\hat{\theta} - g_j\right\}\right\},\ \text{for all } j\right\}.$$

Let $\hat{D} = -\sum_{i=1}^{n} Z_i W_i'/n$.

The level α critical value is the $1-\alpha$ quantile of over bootstrap replications b of

$$T^b = \min_{h\in\hat{V}_n} \hat{Q}^b(\theta) - \min_{h} \hat{Q}^b(\theta),\quad \hat{Q}^b(\theta) := \left\{(W^b + \hat{G}h)'\hat{\Omega}^{-1}\left(W^b + \hat{G}h\right)\right\}^{1/2}.$$

This is the bootstrap version of the difference of constrained and unconstrained objective function.

The use of square root objective function helps in asymptotic theory.

Computation of T^b quite straightforward because the objective function is quadratic and constraints are linear inequalities and or equalities.

$$\hat{V}_n = \left\{ h : Fh = 0,\ G_j h \leq \max\left\{0, -\sqrt{n}\left\{r_n + G_j\hat{\theta} - g_j\right\}\right\},\ \text{for all } j\right\},$$

$$T^b = \min_{h \in \hat{V}_n} \hat{Q}^b(\theta) - \min_h \hat{Q}^b(\theta), \quad \hat{Q}^b(\theta) := \left\{(W^b + \hat{G}h)'\hat{\Omega}^{-1}\left(W^b + \hat{G}h\right)\right\}^{1/2}.$$

Critical value depends on slackness parameter r_n; asymptotics requires r_n converges to zero slower than convergence rate of $\hat{\theta}$, i.e. slower than $1/\sqrt{n}$ for parametric estimator $\hat{\theta}$.

Setting $r_n = +\infty$ always theoretically valid, but may be conservative (all constraints are binding in bootstrap) and result in loss of power.

Heuristically, when r_n tends to zero any constraint that is not binding at θ_0 will also not be binding in the bootstrap with probability approaching one, so inference is not asymptotically conservative for a fixed data generating process.

Example: Testing $H_0 : \theta_0 \le 0$, $\theta_0 = E[Y]$.

The moment function, optimal weighting matrix, GMM objective function and estimator:

$$\begin{aligned} \hat{g}(\theta) &= \sum_{i=1}^{n} (Y_i - \theta)/n = \bar{Y} - \theta, \ \hat{\Omega} = \sum (Y_i - \bar{Y})^2/n, \\ \hat{Q}(\theta) &= \sqrt{n} \left| \bar{Y} - \theta \right| / \hat{\Omega}^{1/2}, \ \hat{\theta} = \min\{0, \bar{Y}\}. \end{aligned}$$

The test statistic is

$$T = \sqrt{n} \left| \max\{\bar{Y}, 0\} \right| / \hat{\Omega}^{1/2}.$$

To describe the critical value, note that $\hat{g}_i = Y_i - \bar{Y}$, $w_i^b \overset{d}{=} N(0,1)$, $W^b = n^{-1/2} \sum_{i=1}^{n} w_i^b (Y_i - \bar{Y})$,

$$\hat{V}_h = \{h : h \le \max\{0, -\sqrt{n}[r_n + \hat{\theta}]\}\}.$$

Critical value is $1 - \alpha$ quantile of bootstrap distribution of

$$T^b = \min_{h \in \hat{V}_h} \{ \left| W^b - h \right| / \hat{\Omega}^{1/2} \}.$$

$$T^b = \min_{h \in \hat{V}_h} \{ \left| W^b - h \right| / \hat{\Omega}^{1/2} \}, \hat{V}_h = \{ h : h \leq \max\{0, -\sqrt{n}[r_n + \hat{\theta}]\} \}.$$

Uniformity comes from behavior of $\hat{V}_h$ as n grows, i.e. behavior of

$$\max\{0, -\sqrt{n}[r_n + \hat{\theta}]\}.$$

Basic assumption is $r_n \longrightarrow 0$, $\sqrt{n} r_n \longrightarrow \infty$.

Suppose true $E[Y]$ is θ_n with $\theta_n \longrightarrow 0$ and $\sqrt{n}\theta_n \longrightarrow -\delta$; this is where exact critical value depends on unknown δ, so do not have uniformity.

Here $\sqrt{n}\hat{\theta} = O_p(1)$, so $-\sqrt{n}[r_n + \hat{\theta}] < 0$ with probability approaching one, so with probability approaching one

$$\hat{V}_h = \{h : h \leq 0\}.$$

T^b is conservative for because it imposes $h \leq 0$ when a weaker (but unknown) restriction would be sufficient for correct critical value from bootstrap.

Smaller choices of r_n give more powerful tests and tighter confidence intervals.

Can choose r_n based on data; intuitively r_n meant to quantify sampling uncertainty in $G(\hat{\theta}-\theta_0)$; cannot estimate distribution of $G(\hat{\theta}-\theta_0)$ uniformly consistently, so link r_n to sampling uncertainty $G(\hat{\theta}^u - \theta_0)$.

Choose γ_n to be small probability giving r_n such that for bootstrap version $\hat{\theta}^{u*}$ of θ_0

$$\Pr(\max_j \{G_j(\hat{\theta}^u - \hat{\theta}^{u*})\} \leq r_n) \approx 1 - \gamma_n;$$

γ_n is more interpretable than r_n; in simulations choice of γ_n does not matter much.

Can get confidence intervals for a linear combination $F\theta_0$ while imposing $G\theta_0 \leq g$ by inverting the test statistic.

For a given α find set of f such that test statistic is less than or equal to critical value for

$$H_0 : F\theta_0 = f,\ G\theta_0 \leq g.$$

Paper describes test and proves validity for partially identified parameters.

Here need additional tuning parameter for estimating identified set that is upper bound on objective function.

Paper describes test and proves validity for nonlinear moment, nonparametric θ models.

Need additional tuning parameter for nonlinear moments involving linearization of moment restrictions.

EMPIRICAL EXAMPLE

We illustrate the preceding discussion by revisiting the study by Angrist and Evans (1998) on the causal effect of childbearing on female labor force participation.

Use the 1980 Census Public Use Micro Sample restricted to mothers aged 21-35 with at least two children.

Outcome of interest is binary variable indicating whether mother is employed.

Treatment binary variable indicating mother has more than two children.

Instrument is indicator for whether the first two children are of the same sex.

Parameter of interest is average treatment effect for compliers (LATE).

Angrist and Evans (1998) document that the impact of childbearing on labor force participation depends on observable characteristics.

In particular, their two stage least squares (2SLS) estimates suggest a negative impact of childbearing on labor force participation across different levels of schooling, with magnitude of the impact decreasing with schooling.

Phenomenon may reflect the fact that more educated mothers have a stronger attachment to the labor force.

To examine this claim we introduce dummy variables S for each year of schooling between 9 and 16 and for the categories less than 9 and more than 16.

We test whether: (i) LATE(s) is increasing in schooling, and (ii) LATE(s) is increasing in schooling and nonpositive.

Both hypotheses fall within the framework of linear IV with inequality restrictions.

LATE(s) is identified through linear moment restrictions and the hypothesized restrictions are linear in LATE(s).

Use five thousand bootstrap replications and setting $r_n = +\infty$ or r_n as determined by $\gamma_n = .05$.

The p-value for LATE(s) being nondecreasing is 0.21.

The p-value for LATE(s) being nondecreasing and nonpositive is 0.394.

Figure 1 gives values of LATE(s) at different schooling levels.

The first panel displays the unconstrained 2SLS estimates and their monotonicity restricted counterparts.

The latter are negative and hence also requiring nonpositive effects does not change the estimates.

Second panel of Figure 1 we give 95% confidence intervals while imposing monotonicity.

Set $\gamma_n = .05$ to get r_n.

Imposing monotonicity yields confidence intervals that are sometimes substantially shorter than 2SLS counterparts.

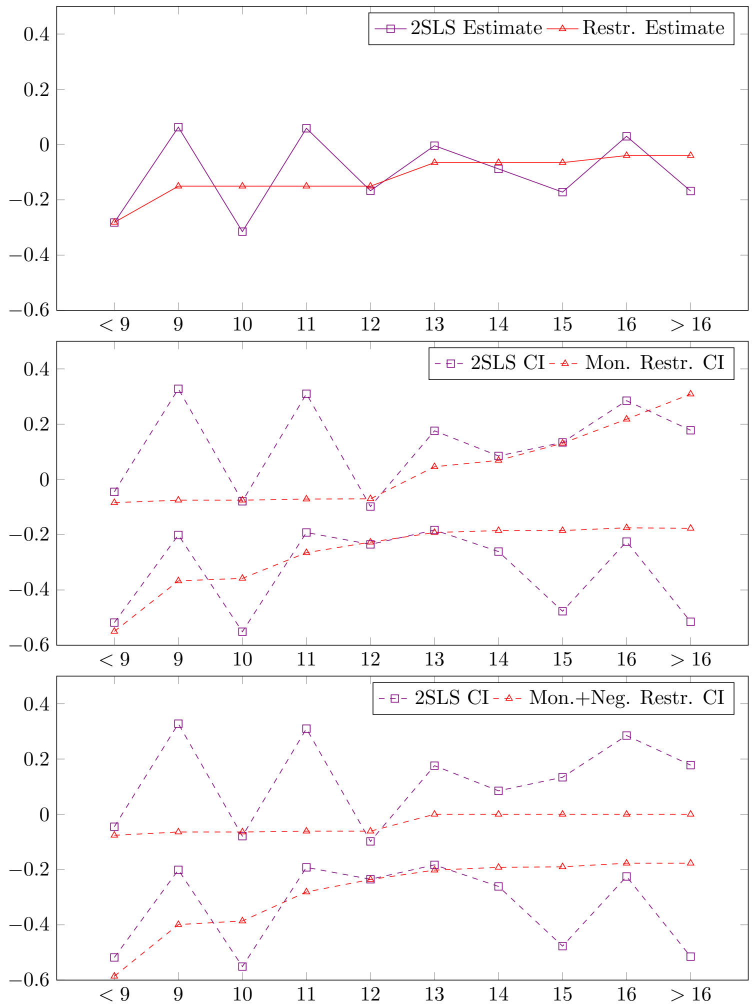

Figure 1: First Panel: Unconstrained and shape restricted LATE estimates (imposing monotonicty or monotonicity and negativity yield the same estimates). Second and Third Panels: 95% Confidence intervals for LATE at different education levels.

to mothers aged 21-35 with at least two children, and set: (i) $D \in \{0, 1\}$ to indicate whether a mother has more than two children (the treatment); (ii) $Y \in \{0, 1\}$ to indicate whether a mother is employed (the outcome of interest); and (iii) $Z \in \{0, 1\}$ to indicate whether the first two children are of the same sex (the instrument). We further adopt the heterogeneous treatment effects model of Imbens and Angrist (1994) and let Y_d denote the potential outcome under treatment status $d \in \{0, 1\}$ and employ "C," "NT," and "AT" to denote compliers, never takers, and always takers.

Angrist and Evans (1998) document that the impact of childbearing on labor force participation depends on observable characteristics. In particular, their two stage least squares (2SLS) estimates suggest a negative impact of childbearing on labor force participation across different levels of schooling, but that the magnitude of the impact decreases with schooling – a phenomenon that may reflect that more educated moth-

PARTIAL IDENTIFICATION

Illustrate for IV with view to inference about ATE in empirical example.

The identified set is

$$\Theta_0 = \{\theta \in \Theta : E[(Y - W'\theta)Z] = 0\}.$$

Test the null hypothesis that the intersection of Θ_0 and the restricted parameter set is nonempty.

Motivated by application take the restricted parameter set to be

$$R = \{\theta : \Upsilon(\theta) = 0,\ G\theta \leq g\},$$

where $\Upsilon(\theta)$ is possibly nonlinear.

For ease of exposition assume that $\min_{\theta \in \Theta} \hat{Q}(\theta) = 0$, i.e. number of instrumental variables is less than or equal to the number of parameters.

Continue to use restricted and unrestricted minima of GMM objective function.

To allow for partial identification we use $\hat{\Omega}$ as before based on a first step $\bar{\theta}$ that is minimum norm minimizer of initial objective function.

For the bootstrap the local parameter space is now changed to account for nonlinear constraints, as

$$\begin{aligned} \hat{V}(\theta, R) &= \{h : \Upsilon(\theta + h/\sqrt{n}) = 0, \\ G_j h &\leq \max\left\{0, -\sqrt{n}\left\{r_n + G_j\theta - g_j\right\}\right\}, \text{ for all } j. \end{aligned}$$

Bootstrap approximation uses an estimator $\hat{\Theta}_n^r$ of $\Theta_0 \cap R$ given by

$$\hat{\Theta}_n = \{\theta \in \Theta_0 \cap R : \hat{Q}(\theta) \leq \inf_{\theta \in \Theta_0 \cap R} \hat{Q}(\theta) + \tau_n\},$$

where $\tau_n \geq 0$ is a bandwidth, making $\hat{\Theta}_n$ is a set of "near minimizers" of $\hat{Q}(\theta)$; positive $\tau_n \longrightarrow 0$ slower than $1/\sqrt{n}$ ensures consistency of $\hat{\Theta}_n$.

$$\hat{\Theta}_n^r = \{\theta \in \Theta_0 \cap R : \hat{Q}(\theta) \le \inf_{\theta \in \Theta_0 \cap R} \hat{Q}(\theta) + \tau_n\},$$

Use step down procedure inspired by Romano and Shaikh (2010) to choose τ_n; see paper.

As before the statistic is

$$T = \min_{\theta \in \Theta_0 \cap R} \hat{Q}(\theta).$$

The bootstrap critical value is more complicated than before; let

$$\begin{aligned} \hat{g}_i(\theta) &= Z_i \varepsilon_i(\theta) - \frac{1}{n} \sum_{j=1}^{n} Z_j \varepsilon_j(\theta), \ \varepsilon_i(\theta) = y_i - W_i'\theta, \\ \hat{W}^b(\theta) &= \frac{1}{n} \sum_{j=1}^{n} w_i^b \hat{g}_i(\theta), \ \hat{Q}^b(\theta, h) = \{[\hat{W}^b(\theta) + \hat{D}h]' \hat{\Sigma}^{-1} [\hat{W}^b(\theta) + \hat{D}h]\}^{1/2}. \end{aligned}$$

Critical value is $1-\alpha$ quantile over bootstrap of $T^b = \min_{\theta \in \hat{\Theta}_n^r, h \in \hat{V}(\theta, R)} \hat{Q}^b(\theta, h)$.

EMPIRICAL EXAMPLE CONTINUED

Inference for conditional ATE denoted ATE(S), where S is schooling; prior emprical work on this example for unconditional ATE is Zhang et al. (2021).

For partial identification use decomposition

$$\begin{aligned} ATE(S) &= LATE(S)P(C|S) + E[Y_1 - Y_0|AT, S]P(AT|S) \\ &\quad + E[Y_1 - Y_0|NT, S]P(NT|S), \end{aligned}$$

where C, AT, NT, are respectively the compliers, always takers, and never takers.

With exception of $E[Y_0|NT, S]$ and $E[Y_1|NT, S]$ all terms can be identified through linear moment restrictions.

With 10 support points for S there are 60 moments and 80 parameters.

Inference on ATE under i) Logical bounds from $Y_d \in \{0, 1\}$; ii) and ATE increasing in S; iii) and ATE is nonpositive.

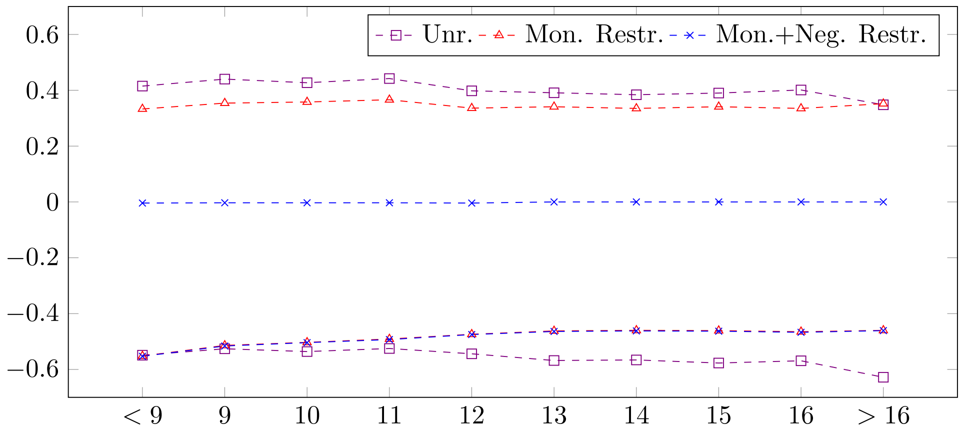

Figure 2: 95% Confidence intervals for ATE at different education levels. "Unr." uses bounds implied by $Y_d \in \{0,1\}$; "Mon. Restr." adds that average treatment effects be increasing in education for all types; "Mon.+Neg. Restr." also requires they be negative.

obtained through the approach described in Remark 2.3 – here, the restriction $G\theta \leq g$ imposes the described shape constraints while the nonlinear restriction $\Upsilon_F(\theta) = 0$ corresponds to imposing a hypothesized value for ATE(S) through (11). In our bootstrap approximation, we let $\tau_n = 0$ and set r_n according to (7) with $\gamma_n = 0.05$ and where we used the distribution of estimators of identified parameters for their partially identified counterparts.[3] We do not report estimates of the identified sets for ATE(S) as they are very close to the obtained confidence intervals: On average the bounds of the confidence intervals exceed the bounds of the estimates by 0.011. Nonetheless, the unrestricted confidence intervals are large as the estimates for the identified set are large – a result driven by the low proportion of compliers (5% on average across S). Imposing monotonicity across types carries identifying information on the upper end of the identified set at low levels of education and on the lower end of the identified set at high levels of education. Additionally imposing nonpositivity sharpens the upper bound of the identified set at all schooling levels. The resulting confidence regions sign ATE(S) at all education levels (weakly) smaller than 12 as strictly negative, though very close to zero.

Finally, as a preview of our general analysis in Section 3, in Table 1 we employ the same shape restrictions to report estimates and 95% confidence intervals for the identified sets of the average treatment effects for: High School Dropouts (edu $\in [9, 12)$), College Dropouts (edu $\in [13, 15)$), College Graduates (edu ≥ 16) and the overall average treatment effect. These confidence regions are obtained through test inversion after noting that a hypothesized value for the average treatment effect of a subgroup can be written as a nonlinear moment restriction in θ_0 through (11) – nonlinear moment restrictions fall within our general framework but outside the scope of Section 2.2. Overall the impact of imposing shape restrictions parallels the results in Figure 2.

[3] E.g., for the constraint $E[Y_1|\texttt{NT}, S] \leq 1$ we substituted the corresponding $G_j\{\hat{\theta}_n^{\text{u}} - \hat{\theta}_n^{\text{u}\star}\}$ term in (7)

Subgroup	Unrestricted		Mon. Restr.		Mon.+Neg Restr.	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
HS Drop	[-0.520,0.426]	[-0.526,0.432]	[-0.489,0.346]	[-0.500,0.356]	[-0.489,-0.008]	[-0.501,-0.003]
Coll. Drop	[-0.561,0.380]	[-0.566,0.385]	[-0.447,0.325]	[-0.460,0.337]	[-0.447,-0.004]	[-0.462,0.000]
Coll. Grad	[-0.579,0.375]	[-0.586,0.382]	[-0.446,0.328]	[-0.462,0.339]	[-0.446,-0.002]	[-0.464,0.000]
All	[-0.545,0.395]	[-0.547,0.398]	[-0.467,0.328]	[-0.477,0.338]	[-0.467,-0.008]	[-0.478,-0.003]

Table 1: Point Estimates and 95% confidence intervals for the average treatment effect at different groups defined by schooling levels under different shape restrictions.

3 General Analysis

We next develop a general inferential framework that encompasses the tests discussed in Section 2. The class of models we consider are those in which the parameter of interest $\theta_0 \in \Theta$ satisfies a finite number $\mathcal{J}$ of conditional moment restrictions

$$E_P[\rho_\jmath(X,\theta_0)|Z_\jmath] = 0 \text{ for } 1 \leq \jmath \leq \mathcal{J}$$

with $\rho_\jmath : \mathbf{X} \times \Theta \to \mathbf{R}$, $X \in \mathbf{X}$, and $Z_\jmath \in \mathbf{Z}_\jmath$. For notational simplicity, we also let $Z \equiv (Z_1,\ldots,Z_{\mathcal{J}})$ and $V \equiv (X,Z)$ with $V \sim P \in \mathbf{P}$. In some of the applications that motivate us, the parameter θ_0 is not identified. We therefore define the identified set

$$\Theta_0 \equiv \{\theta \in \Theta : E_P[\rho_\jmath(X,\theta)|Z_\jmath] = 0 \text{ for } 1 \leq \jmath \leq \mathcal{J}\}$$

and employ it as the basis of our statistical analysis – we emphasize that Θ_0 depends on P, but leave such dependence implicit to simplify notation. For a set R of parameters satisfying a conjectured restriction, we develop a test for the hypothesis

$$H_0 : \Theta_0 \cap R \neq \emptyset \qquad H_1 : \Theta_0 \cap R = \emptyset; \tag{12}$$

i.e. we devise a test of whether at least one element of the identified set satisfies the posited constraint. In what follows, we denote the set of distributions $P \in \mathbf{P}$ satisfying the null hypothesis in (12) by $\mathbf{P}_0$. We also note that in an identified model, a test of (12) is equivalent to a test of whether θ_0 itself satisfies the hypothesized constraint.

The defining elements determining the type of applications encompassed by (12) are the choices of Θ and R. In imposing restrictions on Θ and R we therefore aim to allow for a general framework while simultaneously ensuring enough structure for a fruitful asymptotic analysis. To this end, we require Θ to be a subset of a complete vector space $\mathbf{B}$ with norm $\|\cdot\|_{\mathbf{B}}$ (i.e. $(\mathbf{B},\|\cdot\|_{\mathbf{B}})$ is a Banach space) and consider sets R satisfying

$$R = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}, \tag{13}$$

with a mean zero normal distribution with the variance of the estimator for $E[Y_0|\mathtt{NT}, S]$.

Find wide but not uninformative confidence intervals.

Estimated identified sets quite large due to compliers being a low proportion of population.

Graph gives confidence intervals and tables give confidence sets.

HETEROGENEITY AND DEMAND ANALYSIS

Inference for averages under general heterogeneity in demand.

Let $Y \in [0,1]$ be expenditure share of some commodity, W a vector of prices, total expenditure, and covariates, η indexes preferences.

Here η can be infinite dimensional.

For W independent of η we have for all constants c,

$$\Pr(Y \leq c|W) = \Pr(g(W,\eta) \leq c|W) = \int 1\{g(W,\eta) \leq c\}\mu_0(d\eta).$$

Let $\Psi(g,\eta)$ be an object of interest, such as equivalent variation for some price change and value of W.

Hausman and Newey (2016) consider functionals of the form

$$\int \Psi(g,\eta)\mu_0(d\eta).$$

Can use set up of this paper for inference about such objects, which are generally partially identified.

This set up is random utility model (RUM) of McFadden (2005), Hausman and Newey (2016), Kitamura and Stoye (2018).

We consider a set of η such that for each η the share function $g(W,\eta)$ satisfies Slutzky conditions on a grid of W values.

Can draw randomly from set of share functions rejecting those that do not satisfy Slutzky on a grid to get large set of functions $\{g_s(W)\}_{s=1}^{s_0}$,

$$g_s(W) = \sum_{k=1}^{K} \beta_k^s p_k(W).$$

Specify μ to be discrete with $\Pr(g(W,\eta) = g_s(W)) = \alpha_s \geq 0$.

Allows us to impose smoothness on preferences in the sense that each $g_s(W)$ is smooth.

Leads to residuals of the form

$$\rho_j(Y, W, \theta) = 1(Y \leq c_j) - \sum_{s=1}^{s_0} \alpha_s 1(g_s(W) \leq c_j).$$

Construct moment functions from interactions of residuals with functions of W.

Do GMM and bootstrap imposing the restriction $\alpha_s \geq 0$, $\sum_s \alpha_s = 1$, and $\lambda = \sum_s \alpha_s \Psi(g_s, s)$.

Invert the test statistic to get confidence intervals for λ.

Paper give regularity conditions sufficient for correct level tests and confidence interval coverage.