

Cerquera, Daniel; Laisney, François; Ullrich, Hannes

**Working Paper**

## A note on regressions with interval data on a regressor

DIW Discussion Papers, No. 1419

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Cerquera, Daniel; Laisney, François; Ullrich, Hannes (2014) : A note on regressions with interval data on a regressor, DIW Discussion Papers, No. 1419, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/103368>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Discussion  
Papers

Deutsches Institut für Wirtschaftsforschung

2014

# A Note on Regressions with Interval Data on a Regressor

Daniel Cerquera, François Laisney and Hannes Ullrich

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

#### IMPRESSUM

© DIW Berlin, 2014

DIW Berlin  
German Institute for Economic Research  
Mohrenstr. 58  
10117 Berlin

Tel. +49 (30) 897 89-0  
Fax +49 (30) 897 89-200  
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:  
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:  
<http://ideas.repec.org/s/diw/diwwpp.html>  
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

# A Note on Regressions with Interval Data on a Regressor\*

Daniel Cerquera<sup>†</sup>

François Laisney<sup>‡</sup>

Hannes Ullrich<sup>§</sup>

9 October 2014

## Abstract

Motivated by Manski and Tamer (2002) and especially their partial identification analysis of the regression model where one covariate is only interval-measured, we present two extensions. Manski and Tamer (2002) propose two estimation approaches in this context, focussing on general results. The modified minimum distance (MMD) estimates the true identified set and the modified method of moments (MMM) a superset. Our first contribution is to characterize the true identified set and the superset. Second, we complete and extend the Monte Carlo study of Manski and Tamer (2002). We present benchmark results using the exact functional form for the expectation of the dependent variable conditional on observables to compare with results using its nonparametric estimate, and illustrate the superiority of MMD over MMM.

**Keywords:** partial identification, true identified set, superset, MMD, MMM.

**JEL Classification:** C01, C13, C40.

---

\*The material in this paper is a subset of material circulated in the 2012 ZEW discussion paper “Considerations on Partially Identified Regression Models” written while all authors were affiliated with ZEW, Mannheim. We wish to thank Gordon Klein for useful discussions in the early stages of the development of this work, Chuck Manski and Elie Tamer for their kind encouragements, and Jeff Racine for help in using the np package. We also thank the participants in the DFH Konstanz-Strasbourg Econometrics Seminar and at the 2012 Econometric Society European Meeting in Malaga for their comments. Financial support from the Swiss National Science Foundation (grant PBZHP1-143346) is gratefully acknowledged.

<sup>†</sup>HFU Business School, Furtwangen University, [daniel.cerquera@hs-furtwangen.de](mailto:daniel.cerquera@hs-furtwangen.de)

<sup>‡</sup>BETA, Université de Strasbourg, and ZEW, [laisney@zew.de](mailto:laisney@zew.de)

<sup>§</sup>Department Firms and Markets, DIW Berlin and Department of Business Administration, University of Zurich, [hullrich@diw.de](mailto:hullrich@diw.de)

# 1 Introduction

Weak assumptions or data limitations often lead to incomplete models in econometrics and, as a consequence, to the failure of point identification.<sup>1</sup> In their seminal and well cited contribution, Manski and Tamer (2002), henceforth MT, study identification regions for parameters in regressions with interval data on a regressor or the outcome.<sup>2</sup> In their Monte Carlo experiments, they focus on illustrating the general approach and therefore do not dwell on special aspects of their examples. Here we do the opposite and take advantage of the full knowledge we have of the joint distribution of the regressors to derive exact results.<sup>3</sup> This allows us to complete their study by (i) determining the exact shape of the identified set in each of their examples, (ii) elucidating some aspects of their MC results, and (iii) extending these to an alternative approach they present without illustration. We also compare benchmark results using the exact functional form for the expectation of the dependent variable conditional on observables with results using a nonparametric estimate. The bulk of the rapidly growing literature on the estimation of partially identified models is extremely technical. By contrast we use an ad hoc estimation procedure which is easily understood but also performs sufficiently well in terms of consistency and precision to fulfill our purpose in the limited framework of this paper.

MT consider partial identification of the regression model

$$y = \gamma_1 v + \gamma_2 x + \gamma_3 + \epsilon, \quad E(\epsilon|x, v) = 0, \quad (1)$$

where only the integer interval  $[v_0, v_1]$  to which  $v$  belongs is observed. We focus on the special case with  $v_1 = v_0 + 1$ , which corresponds to their Monte Carlo study. The data generating processes (DGP) MT consider have  $\gamma = (1, -1, 1)$ ,  $\epsilon$  standard normal given  $x$  and

---

<sup>1</sup>See Manski (2003, 2007) for a background on partial identification and Tamer (2010) and Molchanov and Molinari (2014) for recent surveys of applications in econometrics.

<sup>2</sup>By the end of July 2014, Google Scholar reports 256 citations.

<sup>3</sup>This does not mean that we explicitly use that information in the identification strategy, as e.g. Pollmann (2014) suggests. Indeed, we use the criteria proposed by MT to characterize the identified sets, and these depend on this joint distribution, even though its knowledge is not assumed.

$v$  (and thus independent of  $x$  and  $v$ ) while the latter are independent and either normally or uniformly distributed. More precisely,  $x \sim N(1, 4)$ ,  $v \sim N(0, 2)$ , resp.  $x \sim U[0, 5]$ ,  $v \sim U[-2, 3]$ .<sup>4</sup> We refer to these two cases as “the normal case” and “the uniform case”, respectively. For further reference, we need to define functions  $\eta(x, v_0, v_1) = E[y|x, v_0, v_1]$  and  $f(x, v, \gamma) = E[y|x, v]$ .

Note that with the assumptions listed above,  $\eta(x, v_0, v_1) = \gamma_1 \bar{v}(v_0) + \gamma_2 x + \gamma_3$ , with  $\bar{v}(v_0) = E[v|v_0] = v_0 + 1/2$  in the uniform case and

$$\bar{v}(v_0) = \sqrt{2} \frac{\phi(v_0/\sqrt{2}) - \phi(v_1/\sqrt{2})}{\Phi(v_1/\sqrt{2}) - \Phi(v_0/\sqrt{2})} \quad (2)$$

in the normal case.

Our first focus is on obtaining the true identified set,  $C^*$ , for each of these two cases. As discussed in Section 2,  $C^*$  is a polyhedron defined by an infinite set of pairs of inequalities

$$c_1 v_0 + c_2 x + c_3 \leq \eta(x, v_0, v_1) \leq c_1 v_1 + c_2 x + c_3, \quad (3)$$

expressed at all possible arguments of  $\eta$ , where  $(c_1, c_2, c_3)$  is observationally equivalent to the true  $\gamma$ . We show that the model is point identified in the normal case, while in the uniform case the four pairs of inequalities expressed at the four combinations of the bounds for  $x$  and  $v_0$  suffice to characterize  $C^*$ . The solution is found by obtaining the intersections of all triples of planes in  $\mathbb{R}^3$  defined by these eight inequalities, and taking the convex hull of the intersections which satisfy all inequalities.<sup>5</sup> In a similar way, we characterize a (sequence of) superset(s) based on the choice of a (sequence of) moment(s), as will be explained in Section 4.

The second focus of the paper is on revisiting and extending the Monte Carlo study of

---

<sup>4</sup>The second parameter in the normal distribution stands for the variance, as in MT (on p. 533 they state: “The intervals  $[v_0, v_1]$  [...] have width one. This is  $1/\sqrt{2}$  of the standard deviation of  $v$  under the normal design [...]).

<sup>5</sup>Faster algorithms for vertex enumeration for a polyhedron exist, see for instance Fukuda et al. (1997).

MT in the uniform case. We do not deal with the normal case because there the model is point identified. MT are almost silent on the practical details concerning their “modified minimum distance” (MMD) estimation of the identified set. They merely state (p. 533) “The MMD estimates are obtained using the method of simulated annealing.” While the literature now offers sophisticated estimation procedures with well documented asymptotic properties (see e.g. Chernozhukov et al. (2007) and Bontemps et al. (2012)), here we use the same simple geometric approach in estimation as for computing the true  $C^*$ , using the fact that in the uniform case OLS estimation of the regression of  $y$  on  $x$  and  $v_0$  gives us access to a consistent estimator of  $\gamma$  (the details are given in the beginning of Section 5).<sup>6</sup> We proceed by replacing bounds with min and max over the sample. Manski and Pepper (2000) and Chernozhukov et al. (2013) warn against this in general, but our Monte Carlo study shows that these concerns do not apply in our particular framework.

Using the exact functional form for  $\eta(x, v_0, v_1) = E[y|x, v_0, v_1]$  as a benchmark illustrates the gain of having this information over having to resort to nonparametric estimation. The latter still yields very good results, at least in the larger samples, and this is important as the exact functional form of  $\eta$  is in general unknown. In the benchmark, some characteristics of  $C^*$  are estimated very fast in terms of sample size. This is the case for the number of vertices and the slopes of the sides, that is, the shapes of the projected sets. In most cases, the geometric approach clearly outperforms MT in estimating the bounds. We also produce results for the binary response model (BRM) based on the latent linear model, and (numerically) document the importance of considering the polyhedron rather than the rectangular parallelepiped corresponding to the Cartesian product of the intervals for each coefficient.<sup>7</sup>

We summarize the general characterization of the identified set in MT in Section 2. Section 3 shows that parameters are point identified at infinity in the normal case, Section

---

<sup>6</sup>Note that the ability to consistently estimate the parameter  $\gamma$  used in the DGP does *not* imply point identification of the latter.

<sup>7</sup>R code for all computations discussed in the paper is available from the authors upon request.

4 discusses the characterization of the true identified set in the uniform case, and Section 5 presents the Monte Carlo evidence.

## 2 Definition of the Sharp Identified Set in MT

In their Proposition 4, calling  $C$  the set of possible parameter constellations, MT show that under conditions that are satisfied here,<sup>8</sup> the set  $C^*$  of parameter values observationally equivalent to the true parameter  $\gamma$  is (in a slightly altered form)  $C^* = \{c \in C : P[\bar{V}(c)] = 1\}$ , with  $\bar{V}(c) = \{(x, v_0, v_1) : f(x, v_0, c) \leq \eta(x, v_0, v_1) \leq f(x, v_1, c)\}$ .<sup>9</sup> Here we have  $f(x, v, c) = c_1v + c_2x + c_3$ .

A simplified presentation of the MMD approach is based on their Lemma 2:  $C^* = \arg \min_{c \in C} Q(c, \eta)$ , with  $Q(c, \eta) = Q_1(c, \eta) + Q_0(c, \eta)$ ,

$$Q_1(c, \eta) = E \{w(f(x, v_1, c), \eta(x, v_0, v_1)) \mathbf{1}[f(x, v_1, c) < \eta(x, v_0, v_1)]\},$$

$$Q_0(c, \eta) = E \{w(f(x, v_0, c), \eta(x, v_0, v_1)) \mathbf{1}[f(x, v_0, c) > \eta(x, v_0, v_1)]\},$$

where the function  $w : \mathbb{R}^2 \rightarrow \mathbb{R}$  is *any* function with properties  $w(s, s) = 0$  and  $w(s, t) > 0$  if  $t \neq s$ . The simplification adopted here consists in specifying  $w$  such that  $w(s, t) = 1$  for all  $(s, t)$ . This does not satisfy the first property but this is innocuous for exact identification because for all  $c$ ,  $P[f(x, v_1, c) = \eta(x, v_0, v_1)] = P[f(x, v_0, c) = \eta(x, v_0, v_1)] = 0$ , due to the continuity of variable  $x$  and functions  $f$  and  $\eta$ .

Clearly,  $P[\bar{V}(c)] = 1$  amounts to  $Q_1(c, \eta) = Q_0(c, \eta) = 0$ , and thus

$$C^* = \{c \in C : [Q_1(c, \eta) = 0] \wedge [Q_0(c, \eta) = 0]\}. \tag{4}$$

---

<sup>8</sup>These conditions (MT, p. 520) are  $P(v_0 \leq v \leq v_1) = 1$ ,  $E(y|x, v)$  exists, is weakly increasing in  $v$ , and coincides with  $E(y|x, v, v_0, v_1)$ . The set  $\bar{V}(c)$  is the complement of the set  $V(c)$  defined in their equation (16).

<sup>9</sup>Note that, as  $\epsilon$  is independent of  $(x, v)$ , its distribution plays no role in the definition of  $C^*$ , as the probability involved in that definition only depends on the joint distribution of  $(x, v, v_0, v_1)$ .



The analogy principle then suggests estimating  $C^*$  by the set

$$C_N = \{c \in C : [Q_{1N}(c, \hat{\eta}_N) = 0] \wedge [Q_{0N}(c, \hat{\eta}_N) = 0]\} \quad (5)$$

where  $\hat{\eta}_N$  denotes a consistent estimator of function  $\eta$ , and

$$Q_{1N}(c, \hat{\eta}_N) = E_N \mathbf{1} [f(x, v_1, c) < \hat{\eta}_N(x, v_0, v_1)], \quad (6)$$

with  $Q_{0N}(c, \hat{\eta}_N)$  defined accordingly, and  $E_N$  the sample mean. Since the *sample* mean of the indicators will only be 0 if each one is, this is equivalent to

$$C_N = \{c \in C : f(x_n, v_{0n}, c) \leq \hat{\eta}_N(x_n, v_{0n}, v_{1n}) \leq f(x_n, v_{1n}, c), \quad n = 1, \dots, N\}.$$

A weakened version which is needed to avoid empty solutions is, given a sequence of positive numbers  $\alpha_N = o(N)$ ,

$$C_N = \left\{ \begin{array}{l} c \in C : f(x_n, v_{0n}, c) \leq \hat{\eta}_N(x_n, v_{0n}, v_{1n}) + \alpha_N, \\ \hat{\eta}_N(x_n, v_{0n}, v_{1n}) - \alpha_N \leq f(x_n, v_{1n}, c), \quad n = 1, \dots, N \end{array} \right\}. \quad (7)$$

MT show that if function  $f$  has the monotone-index property, which is the case in all instances we consider here,  $C^*$  will be convex (corollary to Proposition 4).

Two remarks seem in order before closing this section. First, we would like to stress that our theoretical characterization of the true identified set is identical to the characterization in MT, and that the identified set only depends on  $\gamma$  and on the joint distribution of  $(x, v)$  – given how  $v_0$  and  $v_1$  are defined here. Second, while the lack of smoothness in our version of the MMD criterion would make it difficult to derive the asymptotic distribution of an estimator based on that criterion, this is not our focus here, as it was not in MT either.

### 3 Point Identification at Infinity in the Normal Case

The inequalities (3) which characterize  $\bar{V}(c)$  can be written

$$c_1v_0 + c_2x + c_3 \leq \gamma_1\bar{v}(v_0) + \gamma_2x + \gamma_3 \leq c_1v_0 + c_2x + c_3 + c_1, \quad (8)$$

for all values  $(v_0, x)$  in the support of their distributions, or equivalently, by subtracting  $\gamma_1v_0 + \gamma_2x + \gamma_3$  everywhere and denoting  $d_i = c_i - \gamma_i$ ,

$$d_1v_0 + d_2x + d_3 \leq \gamma_1[\bar{v}(v_0) - v_0] \leq d_1v_0 + d_2x + d_3 + c_1. \quad (9)$$

**Proposition 1. Point identification of the coefficient of  $x$ .** Whatever the values of  $v_0$ ,  $d_1$  and  $d_3$  are, as the support of  $x$  is  $\mathbb{R}$ , if  $d_2 \neq 0$ , there will be a set of values of  $x$  large enough in absolute value to violate one of the inequalities and this set will have positive probability, contradicting  $P[\bar{V}(c)] = 1$ . Thus  $\gamma_2$  is point identified.<sup>10</sup>

**Proposition 2. Point identification of the coefficient of  $v$ .** As  $v_0$  is the largest element of  $\mathbb{Z}$  below  $v$ , the support of  $v_0$  is  $\mathbb{Z}$  and by the same argument as in Proposition 1  $\gamma_1$  is identified. Note that this extends also to the case where  $f$  has the single index property of Proposition 4 of MT (p. 544): in point (b) of their proof of the corollary to Proposition 4, we have, with our notation  $P[V(c)] \geq P[-d_2x < d_3 + d_1v_0 - \gamma_1]$ , or, since  $d_2 = 0$  if  $c$  is in  $C^*$ ,  $P[V(c)] \geq P[-d_1v_0 < d_3 - \gamma_1]$ , and this probability is strictly positive if  $d_1 \neq 0$ .

**Proposition 3. Point identification of the coefficient of the constant.** Given that  $\gamma_1$  and  $\gamma_2$  are point identified, the inequalities become  $d_3 \leq \gamma_1[\bar{v}(v_0) - v_0] \leq d_3 + \gamma_1$ , or equivalently:

$$\gamma_1[\bar{v}(v_0) - v_0 - 1] \leq d_3 \leq \gamma_1[\bar{v}(v_0) - v_0]. \quad (10)$$

---

<sup>10</sup>This result holds more generally, as shown in point (b) of the Corollary to Proposition 4 in MT.

Since inequality (10) holds for all values of  $v_0$ , it implies

$$\gamma_1 \sup_{v_0 \in \mathbb{Z}} [\bar{v}(v_0) - v_0 - 1] \leq d_3 \leq \gamma_1 \inf_{v_0 \in \mathbb{Z}} [\bar{v}(v_0) - v_0].$$

Restricting attention to unimodal distributions symmetric around 0, Figure 1 plots standard normal, Cauchy and logistic conditional densities of  $v$  given  $v$  in  $[v_0, v_0 + 1]$ . It shows that, while in the Cauchy case the density becomes flat as  $v_0$  goes to  $-\infty$ , in the normal case the distribution places more weight on values near  $v_0 + 1$  for  $v_0 < 0$  and by symmetry near  $v_0$  for  $v_0 > 0$ , so that:<sup>11</sup>

$$\sup_{v_0 \in \mathbb{Z}} [\bar{v}(v_0) - v_0 - 1] = \inf_{v_0 \in \mathbb{Z}} [\bar{v}(v_0) - v_0] = 0. \quad (11)$$

Thus in the normal case, and for any other distribution satisfying (11) (but we have not attempted to characterize the set of these distributions),  $d_3 = 0$ , so that  $\gamma$  is point identified, even though  $v$  is interval-measured.

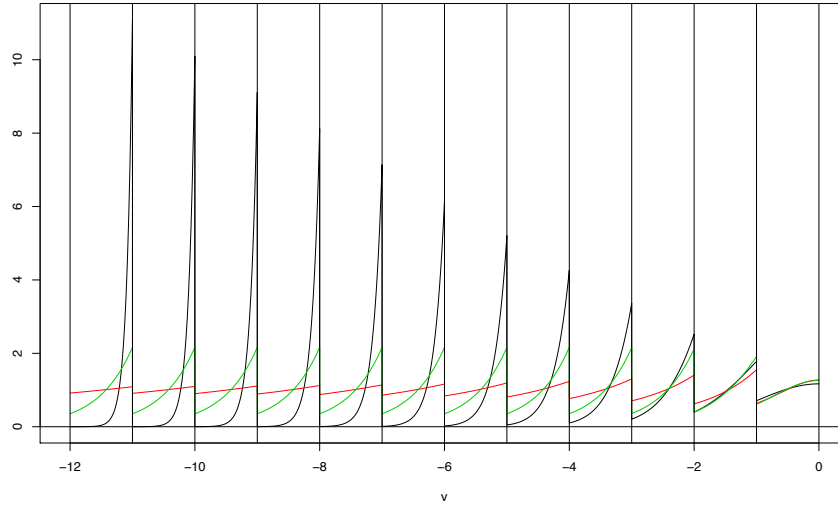


Figure 1: Standard normal, scaled logistic and Cauchy conditional densities on intervals of length 1

<sup>11</sup>Formal proof: let  $t_0 = v_0/\sqrt{2}$ . Then  $\bar{v}(v_0) = \sqrt{2} [\phi(t_0) - \phi(t_0 + 1/\sqrt{2})] / [\Phi(t_0 + 1/\sqrt{2}) - \Phi(t_0)]$ . When  $v_0 \rightarrow +\infty$  the leading term in the numerator is  $t_0\phi(t_0)$ , while in the denominator it is  $\phi(t_0)/\sqrt{2}$ , so that the limit of  $\bar{v}(v_0)$  is then  $t_0\sqrt{2} = v_0$ .

## 4 Set Identification in the Uniform Case

Now the middle term in (9) is  $\gamma_1/2$ . Thus for  $d$  given an upper bound for  $d_1v_0 + d_2x + d_3$  over the values of  $(v_0, x)$  is obtained for the extreme values, depending on the signs of  $d_1$  and  $d_2$ . Then the vertices of the  $C^*$  polyhedron are found as intersections of three of the eight planes in  $\mathbb{R}^3$  corresponding to the two inequalities in (9) written for each of the four pairs combining min and max of  $x$  and  $v$ . For the uniform DGP in MT, the vertices of  $C^*$  are

$c_1$	$c_2$	$c_3$
1.000	-1.000	0.500
1.000	-0.800	0.500
1.000	-1.000	1.500
1.000	-1.200	1.500
0.800	-1.000	1.100
1.333	-1.000	0.833

and this completely characterizes the set  $C^*$ . Figures 2a and 2b show the true polyhedron for this case, and its projections on the planes  $(c_1, c_2)$ ,  $(c_1, c_3)$ , and  $(c_2, c_3)$ .

MT propose another approach, MMM, to estimate a superset but do not provide MC results for illustration. MMM is based on the fact that the inequality  $f(x, v_0, c) \leq \eta(x, v_0, v_1) = E[y|x, v_0, v_1]$  is equivalent with  $E[y - f(x, v_0, c) | x, v_0, v_1] \geq 0$ , which in turns implies that for any vector of  $H$  positive functions  $w(x, v_0, v_1)$  (the arguments of  $w$  are omitted in the sequel),  $E[w\{y - f(x, v_0, c)\}] \geq 0$ . Every point in  $C^*$  thus satisfies both this inequality and  $E[w\{y - f(x, v_1, c)\}] \leq 0$ , and by choosing a specific vector function  $w$ , MT define a set  $C_w^{**}$  which contains  $C^*$ :

$$C_w^{**} = \{c \in C : E[w\{y - f(x, v_1, c)\}] \leq 0 \leq E[w\{y - f(x, v_0, c)\}]\}.$$

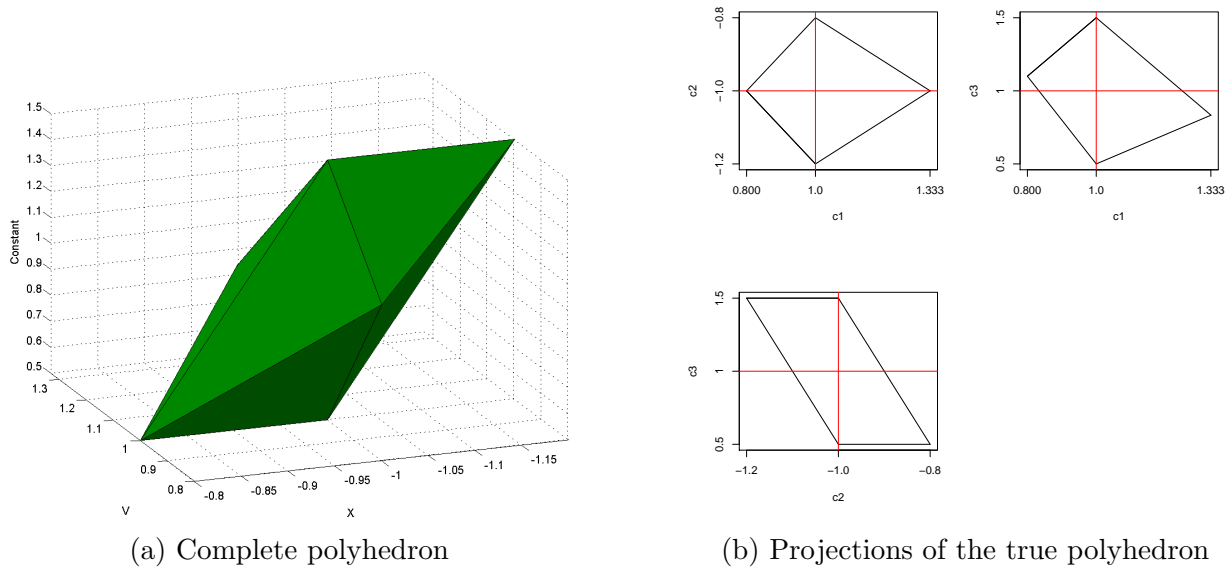


Figure 2: True polyhedron

With  $f(x, v, c) = c_1v + c_2x + c_3$ ,  $C_w^{**}$  is thus defined by a system of  $2H$  linear inequalities involving moments

$$\begin{aligned}
 E[wy] &\leq c_1E[wv_1] + c_2E[w x] + c_3E[w], \\
 E[wy] &\geq c_1E[wv_0] + c_2E[w x] + c_3E[w].
 \end{aligned}
 \tag{12}$$

Each additional row in  $w$  gives two more inequalities, which leads to a reduction of the set. In this way one can construct a decreasing sequence of sets  $C_w^{**}$ , all admitting  $C^*$  as a subset. Again, the analogy principle leads to characterizing  $C_{wN}$  as the set of points  $c$  satisfying

$$\begin{aligned}
 E_N[wy] - \epsilon_N &\leq c_1E_N[wv_1] + c_2E_N[w x] + c_3E_N[w], \\
 E_N[wy] + \epsilon_N &\geq c_1E_N[wv_0] + c_2E_N[w x] + c_3E_N[w],
 \end{aligned}$$

given a sequence of positive numbers  $\epsilon_N = o(N)$ . An advantage of MMM over MMD is that it does not require estimating  $\eta$ . The obvious drawback is that the set it estimates is larger

than  $C^*$  and that its definition depends on the choice of  $w$ .<sup>12</sup>

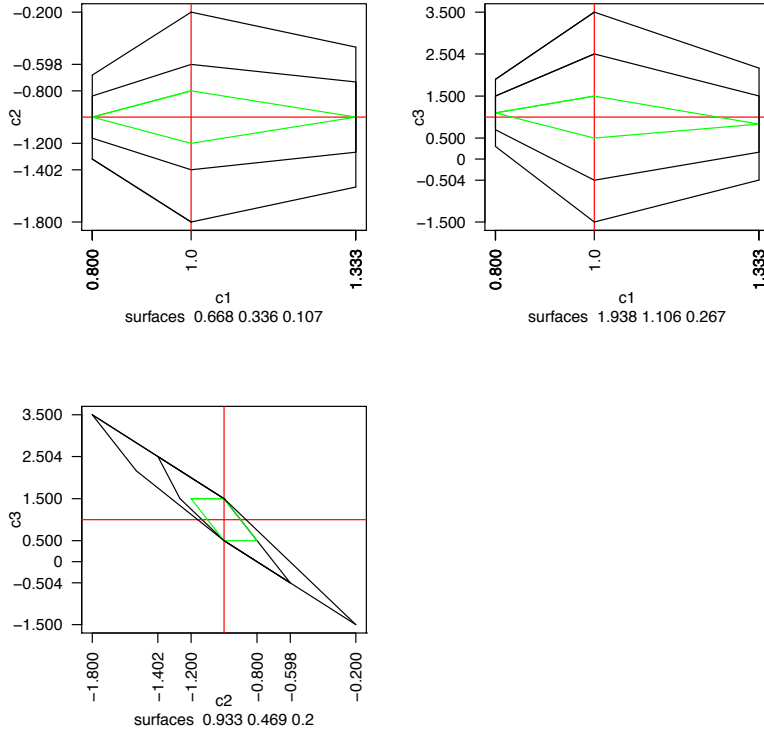


Figure 3: Projections of  $C_w^{**}$  with 2 (resp. 500) powers of  $x/E(x)$

We can compute the exact inequalities corresponding to MMM in the uniform case and the details are given in Appendix A. We approximate the limit of a sequence of  $C_w^{**}$  sets, where the list of “instruments”  $w(x, v_0, v_1)$  contains a complete set of indicators for all possible values of  $v_0$  and powers of  $x/E(x)$ . Figure 3 reports projections of the set for 2 and 500 powers. We were surprised to find that such a large number of moments proved informative, but Menzel (2014) documents an important difference between moment equalities and moment inequalities: if an infinity of moment inequalities apply, no finite subset of these inequalities will define the true identified set. Comparing  $C_w^{**}$  with  $C^*$  we obtain the same bounds on the coefficient of  $v$ , slightly larger bounds on the coefficient of  $x$ , and much larger bounds on the intercept, with a width about 3 instead of 1. The volume of  $C_w^{**}$  (.1119)

<sup>12</sup>An interesting open question is then whether and under which conditions there exists an increasing sequence of functions  $w$  such that the limit of  $C_w^{**}$  coincides with  $C^*$  (a place to start would be the case where  $\gamma$  is point identified, that is  $C^* = \{\gamma\}$ ). Andrews and Barwick (2012), Andrews and Soares (2010), and Andrews and Shi (2013) consider inference in settings with moment inequalities using generalized moment selection procedures. However, as already stated, we do not consider inference here.

is more than three times the volume of  $C^*$  (.0356), as reported in Tables B.1 and B.2 in Appendix B.

## 5 Monte Carlo Evidence for the Uniform Case

We estimate the identified set  $C^*$  with simulated data, using sample analogues and following the geometric approach proposed in Section 4 to characterize the true identified set. We investigate the impact of estimating  $\eta(x, v_0, v_1)$  nonparametrically, as proposed in MT, by comparing the respective results to benchmark results using the exact functional form for  $\eta(x, v_0, v_1)$ . In the sequel, we will refer to the former as ‘nonparametric  $\eta$ ’ and to the latter as ‘exact  $\eta$ ’.

As mentioned in the introduction, we use the same simple geometric approach in estimation as for computing the true  $C^*$ , using the fact that OLS estimation of the regression of  $y$  on  $x$  and  $v_0$  gives us access to a consistent estimator of  $\gamma$  and replacing bounds with min and max over the sample. Indeed

$$\gamma_1 v + \gamma_2 x + \gamma_3 + \epsilon = \gamma_1 v_0 + \gamma_2 x + \gamma_3 + \epsilon + \gamma_1 (v - v_0),$$

and since  $E(v - v_0 | x, v_0) = 1/2$ , the OLS estimator converges to  $(\gamma_1, \gamma_2, \gamma_3 + \gamma_1/2)$  and this allows to retrieve a consistent estimator of  $\gamma$ .

For the exact  $\eta$  case, the complete set of inequalities corresponding to (9) with  $\gamma$  replaced by its estimate and written for all observations in the sample is exactly equivalent, by the argument of the previous section, to the set of eight inequalities corresponding to the maxima and minima of  $(x, v)$ .

For the nonparametric  $\eta$ , this is no longer the case, and by restricting attention to those eight inequalities as a short-cut approximation, we estimate a superset of the set we would obtain using all inequalities. As the MC study shows, this short-cut appears to work quite well. Estimation proceeds in two steps (not counting the OLS step). In the first step,

we estimate  $\eta$  nonparametrically. We do this without using any information on the joint distribution of  $(x, v_0)$  not even independence or bounded support. We use product kernels involving a normal kernel for the continuous variable. For practical reasons we treat  $v_0$  as an ordered factor using a Wang and Van Ryzin kernel (see Hayfield and Racine, 2008).<sup>13</sup> In the second step, the estimated set is found by obtaining the intersections of all triples of planes in  $\mathbb{R}^3$  defined by the two inequalities in (9) written for each of the four pairs combining min and max of  $x$  and  $v$  and taking the convex hull of the intersections which satisfy all inequalities. In principle this last operation should not be needed, since the set  $C^*$  is convex, but it is useful in eliminating spurious vertices resulting from rounding errors.<sup>14</sup> We report results for sample sizes  $N = 100$ ,  $N = 200$ ,  $N = 800$ , and  $N = 20,000$ , as well as the true identified set. In all cases, we use 100 Monte Carlo replications. It might be desirable to consider a larger number of replications. However, here we follow MT and find that our main results show sufficient stability to generate meaningful insights.

In the MC study illustrating MMM, we use the  $2H$  inequalities from Equation (12) to estimate sequences of the sets  $C_w^{**}$ . We replace the moments by their sample analogues and proceed by obtaining the intersections of all triples of planes defined by the inequalities and taking the convex hull of the intersections which satisfy all inequalities. In each sequence, we gradually increase the number of powers of scaled  $|x|$ , denoted by  $k$ , in the function  $w$ . The numbers of powers we report are  $k = 2, 5, 20, 30$ . We further include a set of indicators for all possible values of  $v_0$ .

---

<sup>13</sup>In all cases we use least-squares cross validation (LSCV). In several experiments, computing the bandwidth for every replication is too burdensome computationally. To reduce this burden, we perform cross validation only for the first replication in a small sample design. For subsequent replications in the same experiment, we adjust the bandwidths for the continuous variable  $x$  by taking differences in the sample standard deviations into account. The bandwidth for  $v_0$  found for the first replication is used without adjustment for the other replications. In experiments with large  $N$ , we import the bandwidths from experiments with smaller  $N$  and rescale them by  $(N_0/N)^{1/5}$  for  $x$  and  $(N_0/N)^{2/5}$  for  $v_0$  treated as an ordered factor, where  $N_0$  is the small sample size used to compute the initial bandwidth and  $N$  the large sample size. The estimates with nonparametric  $\eta$  could clearly be improved by performing LSCV for each Monte Carlo replication. See Hayfield and Racine (2008) and the online documentation for the R function `npregbw` in the package `np`.

<sup>14</sup>Some tolerances have to be set here (see Equation (7)), and we chose  $10^{-8}$  for the inequalities, and the default precision for the R function `rankMatrix`, used to check the existence of the intersection of three given planes.



We then investigate the performance of MMM as compared to the MMD method. It is a priori unclear which should perform better in small samples. MT note that we need to weigh the advantage of MMM over MMD, where MMM relies only on a ‘continuous function of unconditional sample moments’ but does not use ‘the full identification power of [MT’s] Proposition 4.’ The latter results in the true  $C^*$  being a subset of any  $C_w^{**}$ .

We construct ‘confidence intervals’ around the means of the estimated bounds such that they include the estimated bounds of 90% of the MC replications.<sup>15</sup> Figures 4 to 6 show selected projections of the true polyhedral set, the overall Minkowski averages, and the Minkowski averages of quantiles based on Hausdorff distances from the true polyhedral set.<sup>16</sup> The fat grey line delineates the true set, the thin grey lines show the estimates, and the solid black line the Minkowski average of all estimated sets. The dotted set is the Minkowski average of estimated sets with the 25% smallest Hausdorff distances to the true set. The dashed and dotted set is the Minkowski average of estimated sets with the 25% largest Hausdorff distances to the true set.

## 5.1 Findings: Modified Minimum Distance

Table 1 reports MT’s MMD results including their confidence regions,<sup>17</sup> and our MMD results for both exact and nonparametric  $\eta$ . We obtain several main findings. Using the exact  $\eta$  as a benchmark illustrates the gain of having this information over having to use its nonparametric estimate. The latter still yields very good results, which is important as the exact functional form of  $\eta$  is a priori unknown to the econometrician. Some characteristics of  $C^*$  are estimated very fast (in terms of sample size). This is the case for the number of vertices and the slopes of the sides, that is, the shapes of the projected sets. In most cases,

<sup>15</sup>The quotation marks emphasize the fact that we merely summarize the results of the Monte Carlo experiments. We leave the issue of inference aside in this paper.

<sup>16</sup>We adapted Matlab code in Beresteanu and Molinari (2008) to R to compute Hausdorff distances and Minkowski averages. For reasons of presentation we have delegated the full set of Figures to an Online Appendix available at [http://hannesullrich.com/mcappendix\\_note.pdf](http://hannesullrich.com/mcappendix_note.pdf)

<sup>17</sup>MT define as confidence region the shortest interval covering 95% of the estimated intervals. Note that this is consistent with our choice of 90% because we interpret their intervals as one-sided.

the geometric approach outperforms MT in estimating the bounds. Finally, we find that in many cases a sample size of 20,000 is not sufficient to approximate the true set. We also produce results for the binary response model based on the latent linear model.

Let us consider these results in more detail. First, the results confirm the expected gain in precision using the exact  $\eta$  over its nonparametric estimate. Even with the smallest sample considered,  $N = 100$ , the mean estimated bounds using the exact  $\eta$  are very close to the true bounds. The absence of pattern in the means for different sample sizes is compatible with absence of bias in the estimation of these bounds. The main benefit from increased sample sizes is tighter confidence intervals over the estimated bounds due to lower sampling imprecision. With nonparametric  $\eta$ , sample sizes of 800 or even 20,000 are needed to approach the true bounds comparatively closely. Now a clear pattern appears, with a decreasing overestimation of the lower bound and a decreasing underestimation of the upper bound, translating into a decreasing underestimation of the width of the identified interval. The bias in the nonparametric estimation of  $\eta$  translates into a bias on the estimation of the bounds. As expected, when the intervals cover the true values, the lower bound increases with  $N$ , and the upper bound decreases. The few exceptions reported in the tables are most probably related to the smallish number of Monte Carlo replications. Table B.1 in Appendix B shows that the volume of the estimated polyhedron approaches the true value of .0356 very fast for exact  $\eta$ . With  $N = 100$ , the average estimated volume is .0362. For nonparametric  $\eta$ , it comes comparatively close with a volume of .0334, using  $N = 20,000$ , but there is substantial underestimation for the smaller sample sizes.

Second, Table 1 shows fast convergence of the MMD estimate to the bounds of the true set. The only instance in which a confidence interval fails to cover the true bound is with a sample size of 100 and nonparametric  $\eta$ , for the upper bound of the  $v$  coefficient. In comparison with the criterion function approach in MT, where the latter requires 20,000 observations for several estimated bounds to approach the true bounds, our estimates mostly achieve this with 800 observations. The lower bound of the coefficient of  $v$  and the bounds

of the intercept even do with as few as 100 observations. Figure 4, and to a lesser extent Figure 5 with nonparametric  $\eta$ , demonstrate how well the MMD estimate picks up the shape of the true set even with very small sample sizes.

Third, compared to MT, the geometric approach is more successful in recovering the bounds of the true set in most cases. This is to some extent expected because we explicitly obtain the vertices of the polyhedron and thus directly find the extreme points of the identified set. The bounds reported by MT for the uniform distribution are too narrow and even their confidence intervals fail to cover the true bounds. We observe the largest differences between MT and the geometric approach in the estimates for the intercept, where MT's mean estimates and confidence intervals fail to cover the true bounds. Hence, MT appear to be successful in finding interior points of  $C^*$  but not its boundary.

We further make the observation that reporting one-dimensional intervals may omit important information. This is increasingly the case with higher dimensions of partially identified parameters. Table B.1 shows that the estimated and true polyhedra cover only about 17% of the volume of the Cartesian product of the corresponding intervals in our experiments. Considering the true shape of the identified set instead of the intervals hence significantly reduces the set of relevant parameter constellations.

In the last group of columns of Table 1, we present results for the binary response model (BRM)

$$y = \mathbf{1} [\gamma_1 v + \gamma_2 x + \gamma_3 + \epsilon > 0],$$

based on the linear latent model studied previously, with the same stochastic assumptions. We now have  $E(y|x, v) = \Phi(\gamma_1 v + \gamma_2 x + \gamma_3)$ , where  $\Phi$  denotes the cdf of the standard normal distribution. Since  $\Phi$  is strictly increasing, MT note that the inequalities  $\Phi(x, v_0, c) \leq \eta(x, v_0, v_1) = E[y|x, v_0, v_1] \leq \Phi(x, v_1, c)$  are equivalent with the (linear in  $c$ ) inequalities  $c_1 v_0 + c_2 x + c_3 \leq \Phi^{-1}[\eta(x, v_0, v_1)] \leq c_1 v_1 + c_2 x + c_3$ .

Here we do not consider the exact functional form of  $\eta(x, v_0, v_1)$  for the BRM since we

have already documented the difference between exact and nonparametric  $\eta$  in the LM.<sup>18</sup>

As do MT, overall we find wider confidence regions for the bounds in the BRM than in the LM. Also, the clear patterns that we found for the LM are now more blurred: the estimates for the BRM are clearly noisier. As in the LM, several of MT's estimated bounds and confidence intervals do not include the true bounds. Our estimated intervals always include the true bounds, except for the  $x$  coefficient with smaller sample sizes and for  $N = 800$  for the lower bound of the intercept (an irregularity that may well be due to the low number of Monte Carlo replications). Note also that here the decreasing lower bounds are necessary to eventually cover the true value. While not as apparent as in the LM, our results seem to converge faster to the true bounds in the BRM than MT's. Comparing the results for  $N = 800$ , the mean estimated bounds for the  $v$  coefficient are much closer using the geometric approach than MT. The same comparisons for the  $x$  coefficient and the intercept yield ambiguous results. Our estimates with  $N = 20,000$  come close to the true bounds, including all of them. However, comparison with MT is difficult as they report only one replication for that sample size.

In discussing the results, we share the apparent presumption of MT that the exact identified set for the BRM model coincides with its pendant for the LM. This seems plausible, as apart from the non-identification of the scale of the latent variable, if  $v$  were observed,  $\gamma$  would be point identified in both models (recall the variance of  $\epsilon$  is 1). This is at odds with the results in Magnac and Maurin (2008), who find exact identification if and only if the distribution of  $v$  given  $x$ ,  $v_0$ , and  $v_1$  is uniform. However, their identification results hinge on the hypothesis of complete variation (Assumption NP.2, p. 838), which is not satisfied in the DGPs used here.

---

<sup>18</sup>In estimating  $\eta$  nonparametrically, we experimented with simple nonparametric regression, ignoring the dichotomous character of  $y$  (R function *npreg* in package *np*), nonparametric conditional density estimation (R function *npcdens*), and the Ichimura and Klein-Spady versions of the semiparametric index model (R function *npindex*), using the same approach as for the LM as regards kernel and bandwidth choices. We obtained plausible results with *npreg* only. The local logit approach in Frölich (2006) might yield better results.

Table 1: Monte Carlo Analysis — MMD — One-dimensional Bounds

Coefficient	Linear Model			Binary Response Model		
	Lower/Upper bound	MT	true $\eta$	np $\eta$	MT	np $\eta$
<i>True: .800</i>						
$v$	N = 100		.795 [ .688, .895]	.822 [ .715, .947]		.778 [ .563, 1.077]
	N = 200	.98 (.87)	.799 [ .743, .863]	.834 [ .738, <b>.949</b> ]	1.05 (.73)	.816 [ .602, 1.060]
	N = 800	.99 (.92)	.802 [ .771, .839]	.812 [ .752, .871]	.95 (.81)	.801 [ .683, .954]
	N = 20000	.97	.800 [ .794, .807]	.813 [ .768, .852]	.85	.851 [ .778, .907]
<i>True: 1.333</i>						
	N = 100		1.325 [ 1.147, 1.491]	1.139 [ .935, 1.310]		1.155 [ .829, 1.451]
	N = 200	1.11 (1.23)	1.332 [ 1.239, 1.438]	1.179 [ .999, <b>1.352</b> ]	1.34 (1.91)	1.171 [ <b>.806</b> , 1.410]
	N = 800	1.07 (1.17)	1.336 [ 1.285, 1.399]	1.243 [ 1.150, 1.334]	1.07 (1.29)	1.273 [ .957, 1.525]
	N = 20000	1.27	1.333 [ 1.323, 1.344]	1.299 [ 1.224, <b>1.366</b> ]	1.14	1.391 [ 1.129, 1.553]
<i>True: -1.200</i>						
$x$	N = 100		-1.192 [ -1.312, -1.079]	-1.135 [ -1.329, -.933]		-.868 [ -1.092, -.597]
	N = 200	-1.03 (-1.14)	-1.211 [ -1.284, -1.137]	-1.143 [ -1.265, -.975]	-1.21 (-1.77)	-.907 [ -1.096, -.704]
	N = 800	-1.01 (-1.08)	-1.197 [ -1.234, -1.152]	-1.172 [ -1.246, -1.097]	-1.01 (-1.18)	-1.044 [ -1.180, -.825]
	N = 20000	-1.14	-1.199 [ -1.210, -1.189]	-1.186 [ -1.226, -1.154]	-1.09	-1.183 [ -1.285, -1.018]
<i>True: -.800</i>						
	N = 100		-.787 [ -.920, -.659]	-.865 [ -1.064, -.661]		-.524 [ -.761, -.247]
	N = 200	-.93 (-.82)	-.807 [ -.889, -.736]	-.854 [ -.998, -.740]	-.96 (-.68)	-.586 [ -.796, -.399]
	N = 800	-.84 (-.87)	-.795 [ -.845, -.749]	-.831 [ -.914, -.769]	-.90 (-.75)	-.649 [ -.779, -.531]
	N = 20000	-.84	-.799 [ -.810, -.790]	-.813 [ -.850, -.775]	-.84	-.751 [ -.818, -.672]
<i>True: .500</i>						
Intercept	N = 100		.474 [ .061, .827]	.610 [ .073, 1.093]		-.256 [ -1.176, .594]
	N = 200	.79 (.48)	.522 [ .301, .752]	.541 [ .134, 1.021]	.83 (.35)	-.114 [ -.990, <b>.624</b> ]
	N = 800	.84 (.61)	.484 [ .372, .617]	.531 [ .329, .747]	.80 (.53)	.052 [ -.479, .386]
	N = 20000	.58	.500 [ .475, .527]	.510 [ .378, .650]	.68	.256 [ -.064, <b>.507</b> ]
<i>True: 1.500</i>						
	N = 100		1.487 [ 1.168, 1.835]	1.491 [ 1.004, 1.976]		.941 [ -.226, 1.735]
	N = 200	1.08 (1.38)	1.533 [ 1.322, 1.739]	1.450 [ 1.042, 1.841]	1.22 (2.09)	.944 [ .016, 1.717]
	N = 800	1.01 (1.22)	1.490 [ 1.382, 1.605]	1.497 [ 1.299, 1.698]	.94 (1.35)	1.316 [ .548, <b>1.721</b> ]
	N = 20000	1.31	1.499 [ 1.476, 1.528]	1.506 [ 1.364, 1.640]	1.24	1.523 [ 1.087, <b>1.801</b> ]

Notes: The MT results show only one bound of the confidence regions because these are defined over the entire estimated interval. We report two sets of ‘confidence intervals’, one for the lower bound and one for the upper bound of the estimated interval. Italicized entries indicate the failure to cover the true value of the corresponding bound. The few exceptions with a lower bound decreasing in  $N$  or an upper bound increasing are shown in bold.

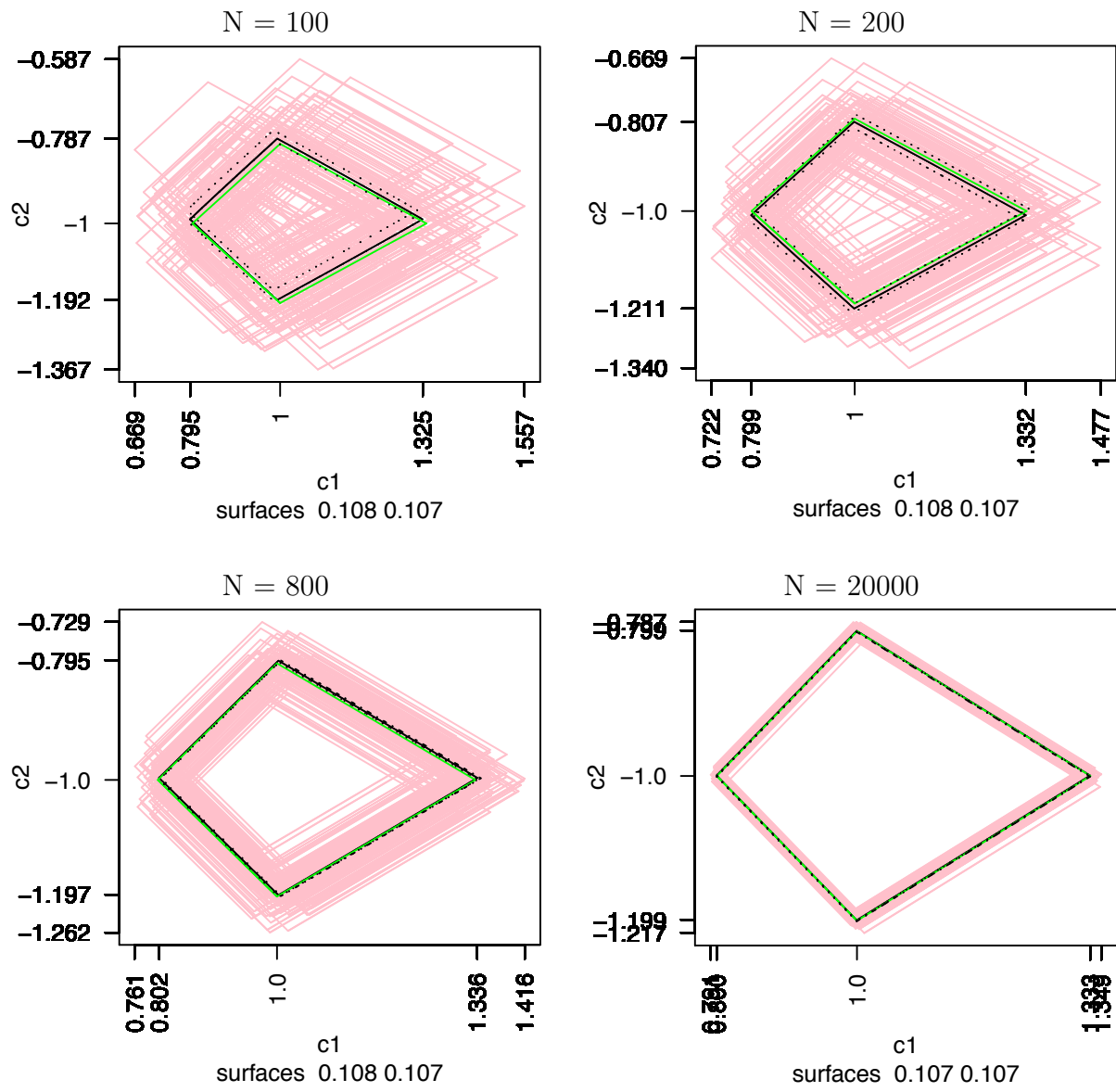


Figure 4: MMD — True  $\eta$  —  $f(x, v, c) = c_1 v + c_2 x + c_3$

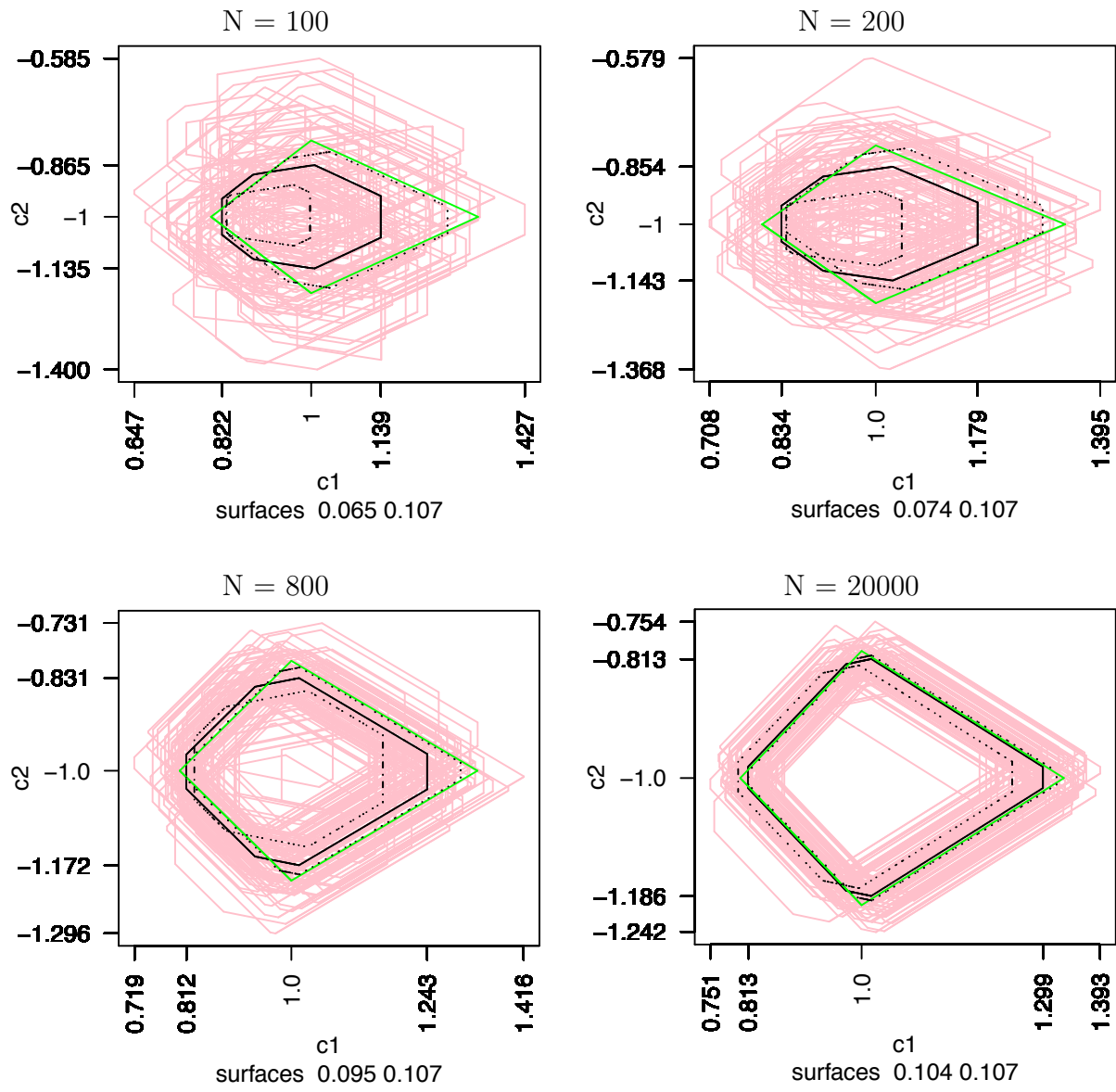


Figure 5: MMD — Nonparametric  $\eta$  —  $f(x, v, c) = c_1 v + c_2 x + c_3$

## 5.2 Findings: Modified Method of Moments

Table 2 reports MMM results for alternative values of  $k$ , the highest power of scaled  $x$  used in  $w$ . We consider  $k = 2, 5, 20, 30$ . As a shortcut we take as reference for  $k = 2$  the values of the true bounds for that value of  $k$ , whereas for the other values of  $k$  considered, the reference bounds are those for  $k = 500$ . This is of no relevance for the coefficient of  $v$ , as the bounds for this are invariant to the choice of  $k$ . The fact that for  $k = 30$  all bounds are covered for all sample sizes supports our choice of not going beyond that value of  $k$ .<sup>19</sup> In general, we find that the convergence of the estimated bounds to the true bounds is improved as the sample size increases. As with MMD, we find that a sample size of  $N = 20,000$  is not always enough to obtain the true bounds. Finally, we find that MMD outperforms MMM, even considering the efficiency cost of estimating  $\eta$  nonparametrically.

More specifically, Table 2 shows the estimation of the bounds. For  $k = 2$ , the estimated bounds for the coefficient on  $v$  converge to the true bounds for  $k = 2$  as the sample size increases. The only case where the true value is not covered is for the upper bound for the intercept and  $N = 100$ . The confidence interval for each estimated bound shows they are precisely estimated with large sample sizes. The setup with  $k = 30$  and  $N = 20,000$  yields tight confidence intervals that contain the true bounds for  $k = 500$ . This is reassuring given the considerable computational cost of estimation with large  $k$ . Overall, the results for  $v$  suggest absence of bias for the bounds for that coefficient. For the other coefficients we find the same tendency to a decreasing overestimation of the identified interval width as with the MMD estimates. Observing Figure 6, which shows projections on the  $(c_1, c_2)$  plane for  $k = 30$ , we see that the shape of the true  $C_w^{**}$  set is not estimated well. Using MMD, the shape of the true set was recovered even with the smallest sample sizes considered, see Figures 4 and 5.

Table B.2 in Appendix B shows that the volume of the estimated superset approaches

---

<sup>19</sup>Limited experiments, not reported here, confirmed that estimated bounds do not vary much as  $k$  is increased to 40 or 50.



and, for large  $N$  and large  $k$ , is nearly identical to, the true value computed with  $k = 500$ : .1119. With  $N = 100$  and  $k = 2$ , the average estimated volume is .1348. As in the discussion of the MMD estimation results, we make the observation in Table B.2 that the volumes of the identified and estimated supersets are only fractions of the product of the corresponding one-dimensional interval lengths. This fraction is below 10% for the MMM estimates which is even smaller than for the MMD estimates. We hence reiterate that considering the true shape of the identified set instead of the intervals significantly reduces the set of relevant parameter constellations.

Table 2: Monte Carlo Analysis — MMM — One-dimensional Bounds

Coefficient	Lower/Upper bound	$k = 2$		$k = 5$		$k = 20$		$k = 30$		
$v$	<i>True: .800, .800</i>									
	N = 100	.806	[.686, .910]	.811	[.688, .910]	.817	[.694, .911]	.820	[.694, .912]	
	N = 200	.795	[.738, .868]	.798	[.744, .870]	.801	[.745, .870]	.801	[.745, .870]	
	N = 800	.794	[.752, .831]	.797	[.754, .835]	.798	[.755, .836]	.799	[.757, .836]	
	N = 20000	.798	[.791, .806]	.799	[.792, .807]	.799	[.792, .807]	.799	[.792, .807]	
	<i>True: 1.333, 1.333</i>									
	N = 100	1.366	[1.149, 1.563]	1.351	[1.146, 1.552]	1.341	[1.146, 1.531]	1.338	[1.146, 1.531]	
	N = 200	1.359	[1.213, 1.492]	1.349	[1.208, 1.481]	1.344	[1.203, 1.473]	1.343	[1.202, 1.473]	
	N = 800	1.363	[1.300, 1.445]	1.355	[1.297, 1.424]	1.351	[1.296, 1.414]	1.351	[1.294, 1.414]	
	N = 20000	1.337	[1.324, 1.351]	1.336	[1.323, 1.350]	1.335	[1.323, 1.349]	1.335	[1.323, 1.349]	
	$x$	<i>True: -1.800, -1.402</i>								
		N = 100	-1.620	[-1.855, -1.409]	-1.449	[-1.637, -1.287]	-1.346	[-1.599, -1.087]	-1.323	[-1.574, -1.008]
N = 200		-1.701	<b>[-1.888, -1.531]</b>	-1.506	[-1.632, -1.375]	-1.402	[-1.557, -1.219]	-1.385	[-1.555, -1.190]	
N = 800		-1.736	[-1.831, -1.633]	-1.521	<i>[-1.608, -1.439]</i>	-1.412	[-1.495, -1.319]	-1.398	[-1.494, -1.292]	
N = 20000		-1.787	[-1.804, -1.767]	-1.552	<i>[-1.565, -1.536]</i>	-1.434	<i>[-1.450, -1.419]</i>	-1.421	[-1.439, -1.402]	
<i>True: -.200, -.598</i>										
N = 100		-.335	[-.582, .009]	-.518	[-.714, -.274]	-.626	[-.864, -.337]	-.647	[-.918, -.337]	
N = 200		-.314	[-.440, -.130]	-.513	[-.633, -.377]	-.619	[-.796, -.462]	-.635	[-.818, -.462]	
N = 800		-.249	[-.335, -.160]	-.468	<i>[-.537, -.403]</i>	-.580	[-.664, -.491]	-.594	[-.696, -.495]	
N = 20000		-.213	[-.236, -.193]	-.447	<i>[-.466, -.432]</i>	-.566	<i>[-.583, -.549]</i>	-.579	[-.600, -.561]	
Intercept		<i>True: -1.500, -.504</i>								
		N = 100	-.963	[-2.110, -.124]	-.522	[-1.476, .106]	-.260	[-1.218, .455]	-.208	[-1.210, .558]
	N = 200	-1.074	[-1.634, -.574]	-.589	[-.986, -.149]	-.332	[-.820, .228]	-.291	[-.807, -.299]	
	N = 800	-1.312	[-1.614, -1.035]	-.774	<b>[-1.041, -.537]</b>	-.497	[-.775, -.231]	-.463	[-.765, <b>-.202]</b>	
	N = 20000	-1.450	[-1.516, -1.377]	-.866	<i>[-.923, -.807]</i>	-.572	<i>[-.629, -.511]</i>	-.539	[-.597, -.472]	
	<i>True: 3.500, 2.504</i>									
	N = 100	2.775	<i>[2.147, 3.438]</i>	2.378	[1.791, 2.950]	2.131	[1.399, 2.802]	2.075	[1.291, 2.798]	
	N = 200	3.077	[2.534, 3.639]	2.604	[2.140, <b>3.012]</b>	2.350	[1.871, <b>2.832]</b>	2.306	[1.836, <b>2.822]</b>	
	N = 800	3.238	[2.913, 3.564]	2.714	[2.487, 2.979]	2.447	[2.175, 2.712]	2.413	[2.137, 2.712]	
	N = 20000	3.449	[3.382, 3.515]	2.864	<i>[2.815, 2.907]</i>	2.570	<i>[2.521, 2.624]</i>	2.538	[2.487, 2.594]	

Notes: We report two sets of ‘confidence intervals’, one for the lower bound and one for the upper bound of the estimated interval. The first set of true bounds corresponds to the true set with  $k = 2$  and the second to  $k = 500$ . Italicized entries indicate the failure to cover the true value of the corresponding bound. For  $k = 2$  the reference is the true bound for  $k = 2$ , for other values of  $k$  we take as reference the true bounds for  $k = 500$ . The few exceptions with a lower bound decreasing in  $N$  or an upper bound increasing are shown in bold.

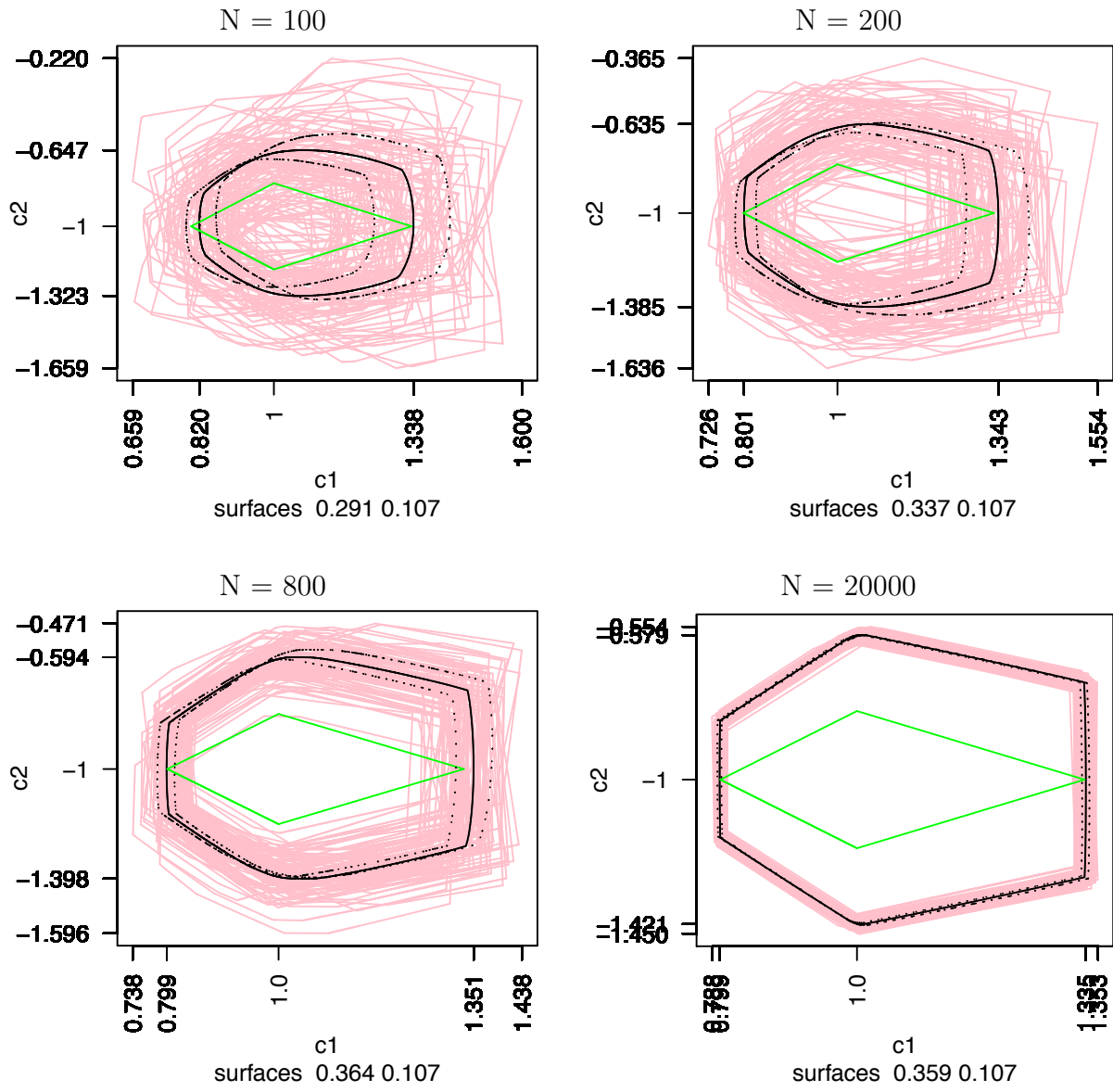


Figure 6: MMM —  $k = 30$  —  $f(x, v, c) = c_1 v + c_2 x + c_3$

## Appendix A True MMM inequalities

We compute the exact form of the following inequalities in the uniform case

$$\Delta_0(c) = E \{w(x, v_0, v_1) [y - f(x, v_0, c)]\} > 0$$

$$\Delta_1(c) = E \{w(x, v_0, v_1) [y - f(x, v_1, c)]\} < 0$$

where

$$f(x, v, c) = c_1 v + c_2 x + c_3,$$

$$y = f(x, v, \gamma) + \epsilon,$$

with  $\gamma_1 = -\gamma_2 = \gamma_3 = 1$ , and  $\epsilon$  independent of all other variables and with zero mean, so that its distribution does not matter for the computation here.  $x \sim U[0, 5]$  and  $v \sim U[-2, 3]$  so that  $v_0$  takes the integer values  $\{-2, -1, 0, 1, 2\}$  with equal probabilities  $1/5$ .

The “instruments”  $w$  are defined as  $w' = (w'_v, w'_x)$ , with  $w_v$  a complete set of indicators for the values of  $v_0$ , and  $w_x$  a vector of the first  $K$  powers of  $x$ :  $w'_x = (x, x^2, \dots, x^K)$ .<sup>20</sup>

Dropping the arguments of  $w$  for simplicity, we can rewrite  $\Delta_j(c)$ ,  $j = 0, 1$ , as

$$\begin{aligned} \Delta_j(c) &= E \{w [f(x, v, \gamma) - f(x, v_j, c)]\} \\ &= \gamma_1 E(wv) - c_1 E(wv_j) + (\gamma_2 - c_2) E(wx) + (\gamma_3 - c_3) E(w). \end{aligned}$$

Note that, because  $x$  and  $v$  are independent,  $E(w_v x) = E(w_v) E(x)$ ,  $E(w_x v) = E(w_x) E(v)$  and  $E(w_x v_j) = E(w_x) E(v_j)$ . Thus all we still need are the eight quantities  $E(v)$ ,  $E(x)$ ,  $E(v_j)$ ,  $E(w_v)$ ,  $E(w_x)$ ,  $E(w_v v)$ ,  $E(w_v v_j)$  and  $E(w_x x)$ .

1.  $E(v) = 5/2 - 2 = 1/2$ .

2.  $E(x) = 5/2$ .

---

<sup>20</sup>Normalizations can be useful to avoid numerical problems: for instance use powers of  $x/E(x)$  or of  $x/\sigma(x)$ , with  $\sigma^2(x) = V(x)$ .

3.  $E(v_0) = 0$ , since  $v_0$  takes the integer values  $\{-2, -1, 0, 1, 2\}$  with equal probabilities  $1/5$ , and  $E(v_1) = 1$ .
4.  $E(w_v)$  : this is a vector with 5 components  $E(1[v_0 = n]) = P[v_0 = n] = 1/5$ .
5.  $E(w_x)$  : this is a vector with  $K$  components  $E(x^k) = (5^k)/(k+1)$ ,  $k = 1, \dots, K$ .
6.  $E(w_v v)$  : this is a vector with 5 components

$$\begin{aligned} E(v1[v_0 = n]) &= E(E(v|v_0 = n)1[v_0 = n]), \quad n = -2, \dots, 2, \\ &= E((v_0 + 1/2)1[v_0 = n]) = (n + 1/2)/5, \end{aligned}$$

since  $E(v|v_0 = n) = n + 1/2$ .

7.  $E(w_v v_0)$  : this is a vector with 5 components

$$E(v_0 1[v_0 = n]) = n/5, \quad n = -2, \dots, 2,$$

and  $E(w_v v_1) = E(w_v (v_0 + 1))$  has components  $(n + 1)/5$ .

8.  $E(w_x x)$  : this is a vector with  $K$  components  $E(x^{k+1}) = (5^{k+1})/(k+2)$ ,  $k = 1, \dots, K$ .

Finally, a choice of  $w_x$  that greatly simplifies computations in the normal case, but that we also may want to try out here to solve numerical problems, is to take a  $K$ -vector of indicators with  $w_x$  the indicator of the  $k$ -th quantile out of  $K$ . Then  $E(w_{xk}) = 1/K$  for all  $k$ , and

$$\begin{aligned} E(w_{xk} x) &= E[w_{xk} E(x|x_{k-1} < x \leq x_k)] \\ &= E[w_{xk} (x_{k-1} + x_k)/2] \\ &= (x_{k-1} + x_k)/(2K). \end{aligned}$$

where  $x_1 = 0$  and  $x_K = 5$ .

**Computational details** The two sides of all inequalities are scaled in order to avoid numerical problems. The solution is found by obtaining the intersections of all triples of planes in  $\mathbb{R}^3$  defined by the  $2H$  inequalities in (12) and taking the convex hull of the intersections which satisfy all inequalities. In principle this last operation should not be needed, since the set  $C_w^{**}$  is convex, but it is useful in eliminating many spurious vertices resulting from rounding errors. Again some tolerances have to be set here, and we chose  $10^{-8}$  for the inequalities, and the default precision for *rankMatrix*.

## Appendix B Monte-Carlo Study: Volumes

Table B.1: Monte Carlo Analysis — MMD — Volumes

	true $\eta$		np $\eta$	
Volume				
<i>True: .0356</i>				
N = 100	.0362	[.0232, .0510]	.0131	[.0004, .0364]
N = 200	.0361	[.0287, .0457]	.0192	[.0011, .0383]
N = 800	.0360	[.0319, .0412]	.0273	[.0124, .0363]
N = 20000	.0355	[.0348, .0365]	.0334	[.0268, .0392]
Volume / Volume “Cube”				
<i>True: .167</i>				
N = 100	.164	[.158, .167]	.170	[.099, .208]
N = 200	.165	[.161, .167]	.172	[.100, .211]
N = 800	.166	[.165, .167]	.186	[.168, .211]
N = 20000	.167	[.167, .167]	.185	[.170, .203]

Table B.2: Monte Carlo Analysis — MMM — Volumes

Volume		
<i>True, k = 2: .2228, k = 500: .1119</i>		
N = 100, k = 2	.1348	[.0465, .2435]
N = 200, k = 5	.1236	[.0587, .1717]
N = 800, k = 20	.1171	[.0980, .1345]
N = 20000, k = 30	.1183	[.1151, .1219]
Volume / Volume “Cube”		
<i>True, k = 2: .052, k = 500: .087</i>		
N = 100, k = 2	.048	[.028, .066]
N = 200, k = 5	.069	[.052, .084]
N = 800, k = 20	.087	[.077, .098]
N = 20000, k = 30	.085	[.083, .088]

## References

- [1] Andrews, Donald W. K. and Panle J. Barwick (2012), “Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure,” *Econometrica*, 80(6), 2805-2826.
- [2] Andrews, Donald W. K. and Xiaoxia Shi (2013), “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 81(2), 609-666.
- [3] Andrews, Donald W. K. and Gustavo Soares (2010), “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78(1), 119-157.
- [4] Beresteanu, Arie and Francesca Molinari (2008), “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76(4), 763-814.
- [5] Bontemps, Christian, Thierry Magnac, and Eric Maurin (2012), “Set Identified Linear Models,” *Econometrica*, 80(3), 1129-1155.
- [6] Cerquera, Daniel, Laisney, François, and Hannes Ullrich (2012), “Considerations on Partial identification of regression models,” Centre for European Economic Research Discussion Paper 2012-024, Mannheim.
- [7] Chernozhukov, Victor, Han Hong, and Elie Tamer (2007), “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75(5), 1243-1284.
- [8] Chernozhukov, Victor, Lee, Sokbae, and Adam M. Rosen (2013), “Intersection Bounds: Estimation and Inference,” *Econometrica*, 81(2), 667-737.
- [9] Frölich, Markus (2006), “Non-parametric Regression for Binary Dependent Variables,” *Econometrics Journal*, 9(3), 511-540.



- [10] Fukuda, Komei, Liebling, Thomas M., and Francois Margot (1997), “Analysis of Back-track Algorithms for Listing All Vertices and All Faces of a Convex Polyhedron,” *Computational Geometry*, 8, 1-12.
- [11] Hayfield, Tristen and Jeffrey S. Racine (2008), “Nonparametric Econometrics: The np Package,” *Journal of Statistical Software*, 27(5).
- [12] Magnac, Thierry and Eric Maurin (2008), “Partial Identification in Binary Models: Discrete Regressors and Interval Data,” *Review of Economic Studies*, 75(3), 835-864.
- [13] Manski, Charles F. (2003), *Partial Identification of Probability Distributions*, Springer, New York.
- [14] Manski, Charles F. (2007), *Identification for Prediction and Decision*, Harvard University Press, Cambridge, MA.
- [15] Manski, Charles F. and John V. Pepper (2000), “Monotone Instrumental Variables: With an Application to the Returns to Schooling,” *Econometrica*, 68(4), 997-1010.
- [16] Manski, Charles F. and Elie Tamer (2002), “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519-546.
- [17] Menzel, Konrad (2014), “Consistent Estimation with Many Moment Inequalities,” *Journal of Econometrics*, 182(2), 329-350.
- [18] Molchanov, Ilya and Francesca Molinari (2014), “Applications of Random Set Theory in Econometrics,” *Annual Review of Economics*, 6, 229-251.
- [19] Pollmann, Daniel (2014), “Identification and Estimation with an Interval-censored Regressor when its Marginal Distribution is Known,” mimeo, Harvard University.
- [20] Tamer, Elie (2010), “Partial Identification in Econometrics,” *Annual Review of Economics*, 2, 167-195.