

Cerquera, Daniel; Laisney, François; Ullrich, Hannes

**Working Paper**

## Considerations on partially identified regression models

ZEW Discussion Papers, No. 12-024

**Provided in Cooperation with:**

ZEW - Leibniz Centre for European Economic Research

*Suggested Citation:* Cerquera, Daniel; Laisney, François; Ullrich, Hannes (2012) : Considerations on partially identified regression models, ZEW Discussion Papers, No. 12-024, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim, <https://nbn-resolving.de/urn:nbn:de:bsz:180-madoc-322108>

This Version is available at:

<https://hdl.handle.net/10419/57183>

**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 Paper No. 12-024

**Considerations on  
Partially Identified Regression Models**

Daniel Cerquera, François Laisney,  
and Hannes Ullrich

**ZEW**

Zentrum für Europäische  
Wirtschaftsforschung GmbH

Centre for European  
Economic Research

Discussion Paper No. 12-024

## **Considerations on Partially Identified Regression Models**

Daniel Cerquera, François Laisney,  
and Hannes Ullrich

Download this ZEW Discussion Paper from our ftp server:

<http://ftp.zew.de/pub/zew-docs/dp/dp12024.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

---

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

## Non-technical summary

In a linear regression model, specifying the conditional mean of a variable (outcome  $y$ ) as a linear function of a set of explanatory variables (regressors  $x$ ),  $E(y|x) = x\theta$ , absence of correlation between the regressors is typically a necessary and sufficient condition for *point identification* of the vector of parameters  $\theta$ , for a wide range of conditional distributions of  $y$  given  $x$ . This means that with sufficiently many observations at hand, a researcher will be able to pinpoint the true value of the parameters.

However, if  $y$  or  $x$  is imperfectly measured, for instance if, for some variable, only interval measurement is available (which is often the case for variables such as age, income, or schooling), then typically point identification will be lost: whatever the number of observations available, there will be a set of values of the parameters that are compatible with the observations. This set is the *identified set*, and the parameter (and also the model) is said to be *partially identified*. The task of the researcher is then to learn from a set of observations what the identified set may be.

The literature on partially identified models has developed considerably over the last ten years and it has many applications in labor economics and more recently in the empirical analysis of game situations in industrial economics. Most of it is extremely technical, and the details needed to apply the estimation methods in practice are often missing. Indeed, this paper originated in an attempt to replicate the illustrative simulation results of the seminal contribution of Manski and Tamer (2002) who study identification regions for parameters in regressions with interval data on a regressor or the outcome.

While their focus was on illustrating the general approach, we concentrate instead on the derivation of exact results for their two Monte Carlo simulation designs. For one of these, the identified set is a simple three-dimensional polyhedron with six vertices and eight faces which is characterized by eight inequalities involving exact expectations. Estimation proceeds by replacing these expectations with sample means. We document significant gains in estimation speed and convergence to the true set, compared with the algorithm used in Manski and Tamer (2002). For the other design, the identified set is more complex, but we show that it can be closely approximated by a simple polyhedron.

## Das Wichtigste in Kürze

In einem linearen Regressionsmodell, in dem der bedingte Erwartungswert einer zu erklärenden Variablen  $y$  als lineare Funktion eines Vektors von erklärenden Variablen  $x$  spezifiziert wird, also  $E(y|x) = x\theta$ , ist die Unkorreliertheit der Regressoren in der Regel, für eine Reihe von bedingte Verteilungen von  $y$  gegeben  $x$ , eine notwendige und hinreichende Bedingung für eine Punktidentifikation des Parametervektors  $\theta$ . Somit können die wahren Parameter anhand einer ausreichend großen Anzahl an Beobachtungen bestimmt werden.

Sind  $y$  oder  $x$  jedoch nicht genau gemessen, sondern, zum Beispiel, nur in Intervallen (häufig sind dies Variablen wie Alter, Einkommen, oder Schulbildung), dann ist eine Punktidentifikation in der Regel nicht mehr möglich: Unabhängig von der Anzahl verfügbarer Beobachtungen existiert eine Parametermenge, die die Beobachtungen erklären kann. Diese Menge ist die *identifizierte Menge*, und ein Parametervektor (genau wie das Modell) wird als *partiell identifiziert* bezeichnet. Die Problemstellung des Forschenden ist damit die Schätzung der identifizierten Menge anhand von Daten.

Die Literatur zu partiell identifizierten Modellen ist in den letzten zehn Jahren stark fortgeschritten und hat insbesondere in der Industrieökonomik viele Anwendungen zur empirischen Analyse spieltheoretischer Modelle gefunden. Ein grosser Teil der Literatur ist äusserst technisch und die Details zur praktischen Anwendung der Schätzmethoden sind häufig nicht dokumentiert. So hat diese Arbeit ihren Ursprung im Versuch die illustrativen Ergebnisse des grundlegenden Beitrags von Manski und Tamer (2002), die die Identifikationsregionen für Parameter in Regressionen mit Intervaldaten über  $y$  oder  $x$  untersuchen, nachzubilden.

Während der Fokus in Manski und Tamer (2002) auf der Illustration des allgemeinen Ansatzes liegt, konzentrieren wir uns auf die Herleitung exakter Ergebnisse aus ihren zwei Monte Carlo Designs. In einem der beiden ist die identifizierte Menge ein einfaches dreidimensionales Polyeder mit sechs Eckpunkten und acht Flächen, das durch acht Ungleichungen, die Erwartungswerte beinhalten, charakterisiert ist. Die Menge kann geschätzt werden, indem wir diese Erwartungswerte durch Stichprobendurchschnitte ersetzen. Wir dokumentieren signifikante Geschwindigkeitsgewinne und eine schnellere Konvergenz hin zur wahren identifizierten Menge im Vergleich zum Algorithmus, den Manski und Tamer (2002) vorschlagen. Für das zweite Monte Carlo Design ist die identifizierte Menge komplexer, aber wir zeigen daß ein simples Polyeder eine sehr gute Annäherung ergibt.

# Considerations on partially identified regression models\*

Daniel Cerquera<sup>†</sup>

François Laisney<sup>‡</sup>

Hannes Ullrich<sup>§</sup>

March 2012

## 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 offer several contributions. 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 estimates, and illustrate the superiority of MMD over MMM. For MMD, we propose a simple shortcut for estimation.

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

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

---

\*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 his np package. We also thank the participants in the DFH Konstanz-Strasbourg Econometrics Seminar for their comments.

<sup>†</sup>Centre for European Economic Research (ZEW) Mannheim, [cerquera@zew.de](mailto:cerquera@zew.de)

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

<sup>§</sup>University of Zurich, [hannes.ullrich@business.uzh.ch](mailto:hannes.ullrich@business.uzh.ch)

# 1 Introduction

Weak assumptions or data limitations often lead to incomplete models in econometrics and, as a consequence, the failure of point identification. Manski (2003, 2007) provides a general background on partial identification and Tamer (2010) a recent survey. In their seminal contribution, Manski and Tamer (2002), henceforth MT, study identification regions for parameters in linear regressions with interval data on a regressor or the outcome.

MT consider two estimation methods, one termed modified minimum distance method (MMD) for estimating the true identified set, and one termed modified method of moments (MMM) for estimating a superset, which will be defined in Section 2. This superset depends on the choice of a sequence of moments. The focus of MT is on illustrating the general approach and therefore they 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. Our first contribution is to characterize the true identified set and the superset. Second, we complete and extend the Monte Carlo study of MT. 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 estimates. In contrast with MT, we also provide illustrations concerning MMM to give helpful insights to applied researchers using data warranting such an approach. A striking feature of the bulk of the literature on partial identification is that the practical details concerning estimation of the identified set are mostly omitted. MT only state (p. 533) “The MMD estimates are obtained using the method of simulated annealing.” Bontemps et al. (2011) show how to obtain a point of the boundary of the identified set in any one direction, but do not discuss how to obtain the complete set. The practical examples they present only consider one direction. Here we emphasize obtaining the entire set. For MMD, we propose a simple shortcut for estimation.<sup>1</sup>

More specifically, we consider the partial identification of the regression model in MT,

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

where only the integer interval  $[v_0, v_1]$  to which  $v$  belongs is observed, and focus on the special case with  $v_1 = v_0 + 1$ , which corresponds to their Monte Carlo study. Without loss of generality, we also assume  $\gamma_1 > 0$ . For further reference, we need to define functions  $\eta(x, v_0, v_1) = E[y|x, v_0, v_1]$  and  $f(x, v, \gamma) = E[y|x, v]$ . The data generating processes (DGP) MT consider have  $\gamma = (1, -1, 1)$ .

Our first focus is on obtaining the true identified set,  $C^*$ , for each of the two DGPs MT consider in their MC study, that is  $x$  and  $v$  are independent and both either uniformly or normally distributed. As we shall see in Section 2,  $C^*$  is a polyhedron defined by an infinite number of inequalities

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

expressed at all possible arguments of  $\eta$ , where  $(c_1, c_2, c_3)$  is observationally equivalent to the true  $\gamma$ . In the *uniform case*, we show that 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

---

<sup>1</sup>R code for all computations discussed in the paper is available from the authors on request.

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>2</sup> In the *normal case*, we show that the model is mathematically point identified at infinity. Yet, using the bisection method with the MMD criterion to obtain  $C^*$ , while confirming identification of the coefficient of  $x$ , leads to an interval for the coefficient of  $v$  that is similar to the one MT report for  $N = 20,000$  in their MC study. It also leads to an interval for the intercept that has almost width 1, as in the uniform case, whereas MT report a much smaller interval. This discrepancy between mathematical identification and what is feasible on a computer is driven by the fact that, while  $v_0$  is potentially unbounded, large absolute values occur with tiny probabilities, so that when computing the minimand of the MMD procedure, only fairly small values of  $|v_0|$  can be used. We also document that a good approximation to this computed solution can be obtained in the same way as for the uniform case, by using sufficiently large values of  $x$  and  $v_0$  (in absolute value) as bounds at which the inequalities are computed. In the sequel, we refer to this as the “polyhedral approximation.”

Although we have only discussed intervals above, we do document the importance of considering the polyhedron rather than the rectangular parallelepiped corresponding to the Cartesian product of the intervals for each coefficient.

Our second focus is on approximating the limit of a sequence of supersets,  $C_w^{**}$ , where the list of “instruments,”  $w$ , includes a complete set of indicators for the values of  $v_0$  (all possible values in the uniform case, the restricted set mentioned above in the normal case) and powers of a scaled version of  $x$ . The solution is found again by obtaining the intersections of all triples of planes in  $\mathbb{R}^3$  defined by the resulting set of moment inequalities, and taking the convex hull of the intersections which satisfy all inequalities. We report sets for 500 powers for the uniform case and, for numerical reasons, only 200 powers for the normal case. We were surprised to find that such a large number of moments proved informative, but Menzel (2011) 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. For the uniform case, and comparing  $C_w^{**}$  with  $C^*$ , we find 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  $C_w^{**}$  is also more than three times the volume of  $C^*$ . For the normal case, we find that  $C_w^{**}$  is very close to our numerically computed  $C^*$ . This is remarkable, because obtaining  $C_w^{**}$  numerically does not require any shrinking of the support of  $v_0$ , contrary to obtaining  $C^*$  numerically.

The third focus of the paper is on *estimation* of  $C^*$  and  $C_w^{**}$ . In both cases we use the same geometrical approach as for computing the true  $C^*$  and  $C_w^{**}$ , replacing expectations with sample means and bounds with min and max over the sample. In the Monte Carlo study, we obtain several main findings. For MMD, 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, which is important as the functional form of the exact  $\eta$  is in general unknown. In the uniform case, some characteristics of  $C^*$  are estimated very fast. This is the case for the number of vertices and the slopes of the sides, that is, the shapes of

---

<sup>2</sup>There exist more refined algorithms for vertex enumeration for a polyhedron, but Fukuda et al. (1997) show that the problem of vertex enumeration of a polyhedron is NP-complete.



the projected sets. The polyhedral approximation performs well in the normal case. In most cases, the geometrical 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 (BRM) based on the latent linear model. 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 linear model. 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 (with the variance of  $\epsilon$  set to 1). This is at odds with the results found 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 an assumption of “complete variation” (Assumption NP.2, p. 838), which is not satisfied in the DGPs used here.

For MMM with alternative values of the powers of scaled  $x$ , we find that the convergence of the estimated bounds to the true bounds is improved as the sample size increases but, as in the MMD case, we find that a large sample size of  $N = 20,000$  is not always enough to achieve convergence to the true bounds. These results are more clear in the uniform case. Finally, we find that MMD outperforms MMM, even considering the cost of estimating  $\eta$  nonparametrically.

We summarize the identification results in MT in Section 2. Section 3 discusses the characterization of the true identified sets and Section 4 presents the Monte Carlo evidence.

## 2 Definition of the sharp identified sets in MT

We begin with a quick exposition of our interpretation of the MMD and MMM approaches in MT. In their Proposition 4, calling  $C$  the set of possible parameter constellations, MT show that under conditions that are satisfied here, the set  $C^*$  of parameter values observationally equivalent to the true parameter  $\gamma$  is (in a slightly altered form):<sup>3</sup>

$$C^* = \{c \in C : P[\bar{V}(c)] = 1\}, \quad (1)$$

with

$$\bar{V}(c) = \{(x, v_0, v_1) : f(x, v_0, c) \leq \eta(x, v_0, v_1) \leq f(x, v_1, c)\}. \quad (2)$$

Here we have

$$f(x, v, c) = c_1 v + c_2 x + c_3. \quad (3)$$

A simplified presentation of the MMD approach is based on their Lemma 2:

$$C^* = \arg \min_{c \in C} Q(c, \eta),$$

with

$$Q(c, \eta) = Q_1(c, \eta) + Q_0(c, \eta),$$

---

<sup>3</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 defined in their equation (16).

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

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

where the function  $w : \mathbb{R}^2 \rightarrow \mathbb{R}$  has 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 is innocuous for exact identification because for all  $c$

$$P[f(x, v_1, c) = \eta(x, v_0, v_1)] = P[f(x, v_0, c) = \eta(x, v_0, v_1)] = 0.$$

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

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

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

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

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

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

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  $\epsilon_N = o(N)$ ,

$$C_N = \left\{ \begin{array}{l} c \in C : f(x_n, v_{0n}, c) \leq \hat{\eta}_N(x_n, v_{0n}, v_{1n}) + \epsilon_N, \\ \hat{\eta}_N(x_n, v_{0n}, v_{1n}) - \epsilon_N \leq f(x_n, v_{1n}, c), \quad n = 1, \dots, N \end{array} \right\}. \quad (5)$$

The other approach, 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 function  $w$ , MT define a set  $C_w^{**}$  which contains  $C^*$ :

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

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

$$\begin{aligned} E[wy] &\leq c_1E[wv_1] + c_2E[w x] + c_3E[w], \\ E[wy] &\geq c_1E[wv_0] + c_2E[w x] + c_3E[w]. \end{aligned} \tag{6}$$

Each additional moment 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>4</sup>

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), and this property extends to  $C_w^{**}$ , as a polyhedron defined by linear inequalities.

### 3 Characterizing the true identified sets

We proceed with a series of remarks that lead to the characterization of  $C^*$  in the special cases MT consider in their MC study. Remark 1 explains why the true parameter  $\gamma$  lies in the identified set  $C^*$ , Remarks 2 and 3 why the coefficients of  $x$  and  $v$  are point identified if the support of  $(x, v)$  is unbounded. Remark 4 concerns identification of the constant. Remark 5 lays out preliminaries for the characterization of  $C^*$ . Remark 6 gives that characterization for the uniform case and Remark 7 justifies the polyhedral approximation for the normal case.

For the moment we do not assume independence between  $x$  and  $v$ , and thus

$$\eta(x, v_0, v_1) = \gamma_1E(v|v \in [v_0, v_0 + 1], x) + \gamma_2x + \gamma_3 = \gamma_1\bar{v}(v_0, x) + \gamma_2x + \gamma_3, \tag{7}$$

where  $\bar{v}(v_0, x)$  denotes  $E(v|v \in [v_0, v_0 + 1], x)$ . The inequalities which characterize  $\bar{V}(c)$  can thus be written

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

---

<sup>4</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  $C^* = \{\gamma\}$ ).

for all values  $(v_0, x)$  in the support of the joint distribution of these variables. Equivalently, by subtracting  $\gamma_1 v_0 + \gamma_2 x + \gamma_3$  everywhere and denoting  $d_i = c_i - \gamma_i$

$$d_1 v_0 + d_2 x + d_3 \leq \gamma_1 [\bar{v}(v_0, x) - v_0] \leq d_1 v_0 + d_2 x + d_3 + c_1. \quad (8)$$

**Remark 1. The true parameter lies in the identified set.** The term  $\bar{v}(v_0, x) - v_0$  lies in  $[0, 1]$ , and thus the middle term lies in  $[0, \gamma_1]$ . Thus, the parameter  $\gamma$  corresponding to the DGP lies in  $C^*$  since for  $c = \gamma$  the inequalities reduce to

$$0 \leq \gamma_1 [\bar{v}(v_0, x) - v_0] \leq \gamma_1,$$

which is satisfied with probability 1.

**Remark 2. Point identification of the coefficient of  $x$ .** Whatever the values of  $v_0$ ,  $c_1$  and  $c_3$  are, if the support of  $x$  is  $\mathbb{R}$  and 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. Thus if  $x$  has unbounded support,  $\gamma_2$  is point identified.

**Remark 3. Point identification of the coefficient of  $v$ .** Still assuming that the support of  $x$  is  $\mathbb{R}$ , assume that the support of  $v$  is also  $\mathbb{R}$ . Then 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 Remark 2,  $\gamma_1$  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:

$$P[V(c)] \geq P[-d_2 x < d_3 + d_1 v_0 - \gamma_1],$$

or, since  $d_2 = 0$  if  $c$  is in  $C^*$ ,

$$P[V(c)] \geq P[-d_1 v_0 < d_3 - \gamma_1],$$

and this probability is strictly positive if  $d_1 \neq 0$ .

**Remark 4. Point identification of the coefficient of the constant.** In the case where both  $x$  and  $v$  are unbounded, and thus  $\gamma_1$  and  $\gamma_2$  are point identified, the inequalities become

$$d_3 \leq \gamma_1 [\bar{v}(v_0, x) - v_0] \leq d_3 + \gamma_1,$$

or equivalently

$$\gamma_1 [\bar{v}(v_0, x) - v_0 - 1] \leq d_3 \leq \gamma_1 [\bar{v}(v_0, x) - v_0]. \quad (9)$$

Assuming independence between  $x$  and  $v$  implies that  $E(v|v \in [v_0, v_0 + 1], x)$  does not depend on

$x$ , and so we denote it  $\bar{v}(v_0)$ . Since inequality (9) 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 3 plots the Cauchy, normal, 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 and more weight on values near  $v_0 + 1$  for  $v_0 < 0$  and by symmetry near  $v_0$  for  $v_0 > 0$ , so that

$$\sup_{v_0 \in \mathbb{Z}} [\bar{v}(v_0) - v_0 - 1] = \inf_{v_0 \in \mathbb{Z}} [\bar{v}(v_0) - v_0] = 0. \quad (10)$$

The logistic case lies inbetween. In the normal case (and for any other distribution satisfying (10)),  $d_3 = 0$ , so that  $\gamma$  is point identified, even though  $v$  is interval-measured.

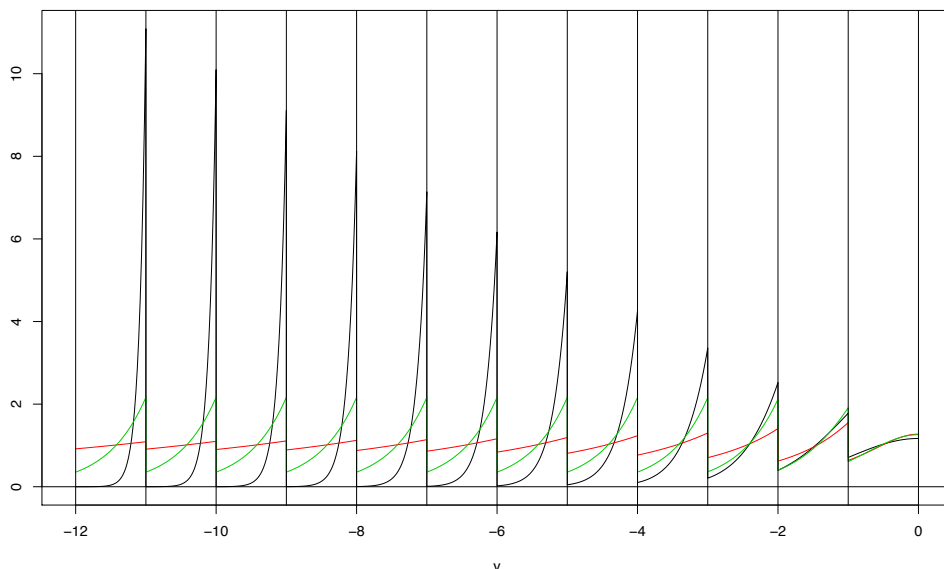


Figure 1: Standard normal, scaled logistic and Cauchy conditional densities on intervals of length 1

**Remark 5. Preliminaries for the characterization of the identified set.** Assume now that  $v$  has bounded support  $[v_l, v_u]$ , and consider the case where  $x$  has unbounded support, so that  $\gamma_2$  is identified. The inequalities in (8) can be rewritten in the more convenient form

$$\begin{aligned} d_1 v_0 + d_3 &\leq \gamma_1 [\bar{v}(v_0) - v_0] \\ \gamma_1 [\bar{v}(v_0) - v_0 - 1] &\leq d_1 (v_0 + 1) + d_3. \end{aligned}$$

Assume for simplicity that  $\bar{v}(v_0) - v_0 = 1/2$  for all values of  $v_0$ , so that now

$$\begin{aligned} d_1 v_0 + d_3 &\leq \gamma_1/2, \\ d_1 (v_0 + 1) + d_3 &\geq -\gamma_1/2. \end{aligned}$$

For a given triple  $(c_1, \gamma_2, c_3)$  in  $C^*$ , bounds of the first terms in these inequalities are found for  $v_0$  equal to its minimum or to its maximum, depending on the sign of  $d_1$ . Thus the vertices of the polygon which is the projection of  $C^*$  on the  $(c_1, c_3)$  plane are found as those intersections of pairs of the four lines

$$\begin{aligned} d_1 v_l + d_3 &= \gamma_1/2 \\ d_1 v_u + d_3 &= \gamma_1/2 \\ d_1 (v_l + 1) + d_3 &= -\gamma_1/2 \\ d_1 (v_u + 1) + d_3 &= -\gamma_1/2 \end{aligned}$$

that satisfy the inequalities. For the pair  $(-12, 11)$  of extreme values for  $v_0$  (there is virtually no probability mass outside that interval if  $v \sim N(0, 2)$ , the situation considered by MT), this yields the following vertices, given first for any  $\gamma$ , then for  $\gamma = (1, -1, 1)$ :

$d_1/\gamma_1$	$d_3/\gamma_1$	$c_1$	$c_3$
0.000	0.500	1.000	1.500
0.000	-0.500	1.000	0.500
-0.043	-0.022	0.957	0.978
0.043	0.022	1.043	1.022

This illustrates how point identification of  $\gamma_1$  is lost when we restrict attention to subset  $[v_l, v_u - 1]$  when the true support of  $v_0$  is  $\mathbb{Z}$ , but we return to a more precise formulation in Remark 7.

**Remark 6. Characterization of the identified set in the uniform case.** Consider now the case where  $x$  has also bounded support  $[x_l, x_u]$ ,  $v$  is independent of  $x$  and uniform, so that the middle term in (8) is  $\gamma_1/2$ . 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 (8) written for each of the four pairs combining min and max of  $x$  and  $v$ . If  $\gamma = (1, -1, 1)$ ,  $v \sim U[-2, 3]$  and the distribution of  $x$  has support  $[0, 5]$  (even if it is not uniform, for instance it could have any other scaled Beta distribution), 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  $C^*$  in the uniform case of MT. Figures 2 and 3 show the true polyhedron for this case, and its projections.

**Remark 7. Polyhedral approximation in the normal case.** Here we introduce and justify the polyhedral approximation we advocate in the normal case. Going back to the framework of Remark

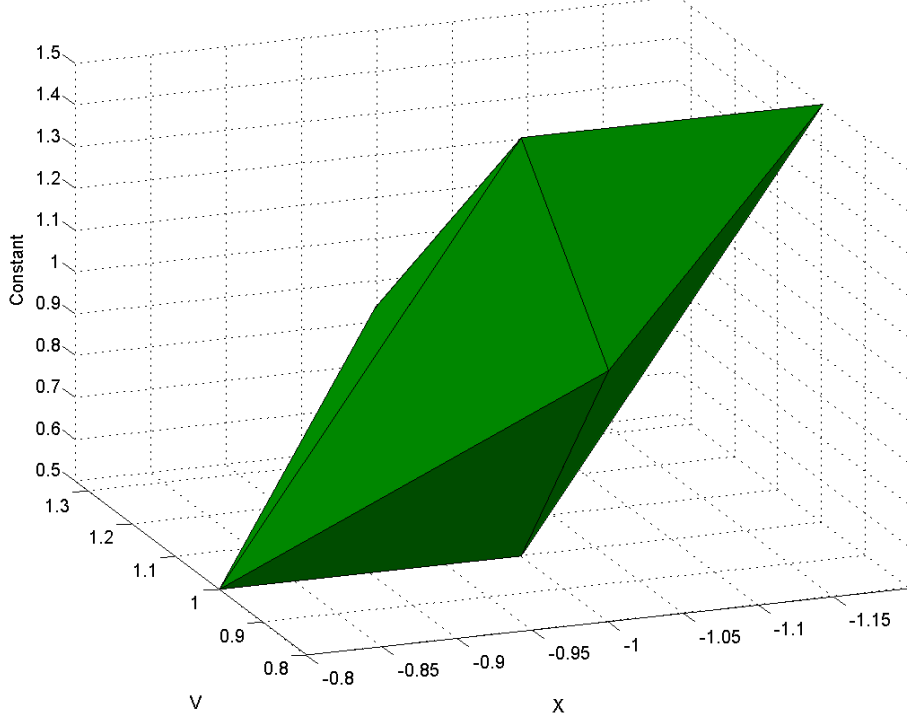


Figure 2: Uniform case: True polyhedron

4, where both  $x$  and  $v$  have unbounded support, assume that, for numerical reasons, we restrict attention to a subset  $[v_l, v_u]$  of the true support of  $v_0$ , which is  $\mathbb{Z}$ . Assume further independence between  $x$  and  $v$  and that the distribution of  $v$  is normal with mean 0: then  $\bar{v}(v_0) - v_0$  is a decreasing function of  $v_0$  (from 1, attained for  $-\infty$ , to 0, attained for  $+\infty$ ). Given identification of  $\gamma_2$ , the inequalities are as in Remark 5

$$\begin{aligned} d_1 v_0 + d_3 &\leq \gamma_1 [\bar{v}(v_0) - v_0] \\ \gamma_1 [\bar{v}(v_0) - v_0 - 1] &\leq d_1 (v_0 + 1) + d_3. \end{aligned} \tag{11}$$

Since

$$\bar{v}(v_u) - v_u \leq \bar{v}(v_0) - v_0 \leq \bar{v}(v_l) - v_l$$

we can relax (resp. strengthen) the first inequality in (11) as

$$d_1 v_0 + d_3 \leq \gamma_1 [\bar{v}(v_l) - v_l] \quad (\text{resp. } \gamma_1 [\bar{v}(v_u) - v_u]).$$

Thus, still denoting  $C^*$  the set defined by inequalities (11) over  $v_0 \in [v_l, v_u]$ , although it is strictly larger than  $C^*$ , we can define two sets  $C_l^*$  and  $C_u^*$  such that  $C_l^* \subset C^* \subset C_u^*$ :  $C_l^*$  is defined by the inequalities

$$\begin{aligned} d_1 v_0 + d_3 &\leq \gamma_1 [\bar{v}(v_u) - v_u] \\ \gamma_1 [\bar{v}(v_l) - v_l - 1] &\leq d_1 (v_0 + 1) + d_3 \end{aligned}$$

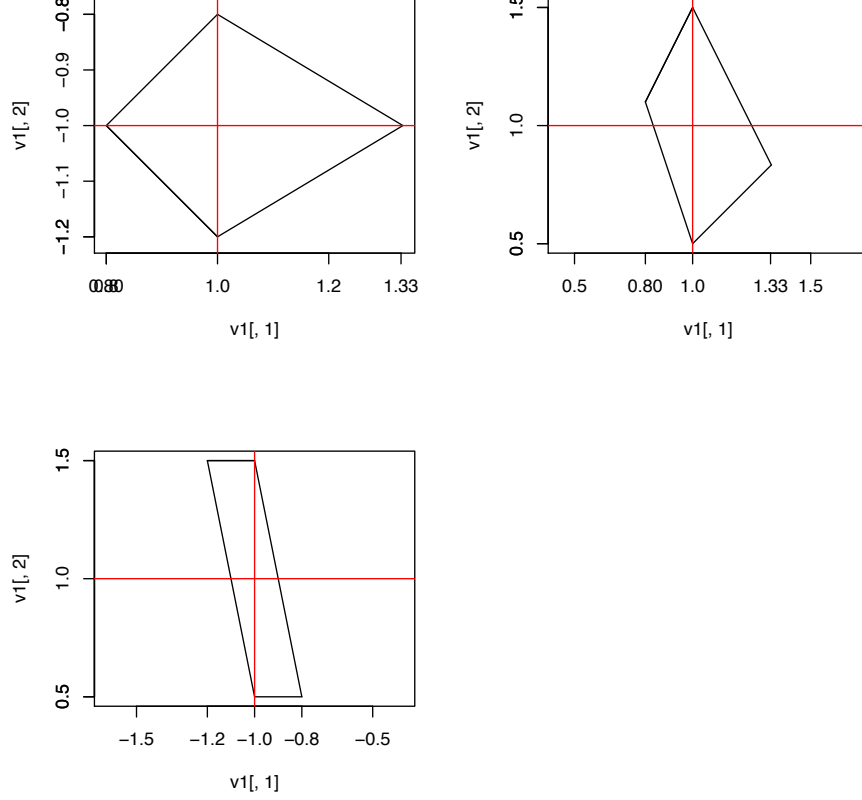


Figure 3: Uniform case: Projections of the true polyhedron

and  $C_u^*$  by the inequalities

$$d_1 v_0 + d_3 \leq \gamma_1 [\bar{v}(v_l) - v_l]$$

$$\gamma_1 [\bar{v}(v_u) - v_u - 1] \leq d_1 (v_0 + 1) + d_3.$$

Both sets are polygons whose vertices can be obtained in the usual way by writing the inequalities above for the pair of values  $(v_l, v_u)$  of  $v_0$ . For  $v \sim N(0, 2)$  and the pair  $(-12, 11)$ , this yields (as in Remark 5, we could have expressed these for any DGP, but for short we give the results for  $\gamma = (1, -1, 1)$ )

$C_l^*$		$C_a^*$		$C_u^*$	
$c_1$	$c_3$	$c_1$	$c_3$	$c_1$	$c_3$
1.000	0.812	0.970	0.485	1.000	1.812
1.000	1.188	0.970	1.485	1.000	0.188
0.982	0.991	0.923	0.961	0.923	0.961
1.018	1.009	1.018	1.009	1.077	1.039

The columns in the middle correspond to the polyhedral approximation  $C_a^*$  based on the inequalities (11) evaluated at  $v_l$  and  $v_u$ . Note that the bounds for  $c_1$  are  $[0.982, 1.018]$  for  $C_l^*$  and  $[0.923, 1.077]$  for  $C_u^*$ , with the expected inclusion, and that the bounds for  $C_a^*$ ,  $[0.923, 1.018]$ , lie in between. For  $c_3$ , the bounds are  $[0.812, 1.188]$ ,  $[0.188, 1.812]$  and  $[0.485, 1.485]$ , again with the expected inclusions, although  $C_l^*$  and  $C_u^*$  are much less informative as regards  $c_3$  as they are concerning  $c_1$ .



In the linear model it is feasible to compute the asymptotic MMD criterion, using analytical expressions for expectations. Here we no longer assume it known in advance that  $\gamma_2$  is identified. We obtain the (numerically) identified set  $C^*$  using a bisection method. The details are given in Appendix A. Figure 4 shows projections of  $C^*$  on the  $(c_1, c_3)$  plane for the range  $[-1.015, -0.985]$  of values of  $c_2$  (it proved numerically difficult to go beyond these values). Figure 5 shows projections of the polyhedral approximation of Figure 4 based on evaluating the inequalities (8) at the four combinations of the pairs of values  $(-12, 11)$  for  $v_0$  and  $\pm 10^6$  for  $x$  (we will still refer to it as  $C_a^*$ ).<sup>5</sup> Other choices that lead to approximations rather than true sets concern various tolerance levels. See Appendix A for details. This said, the polyhedral approximation to the true set appears quite satisfactory, as Figures 4 and 5 show.

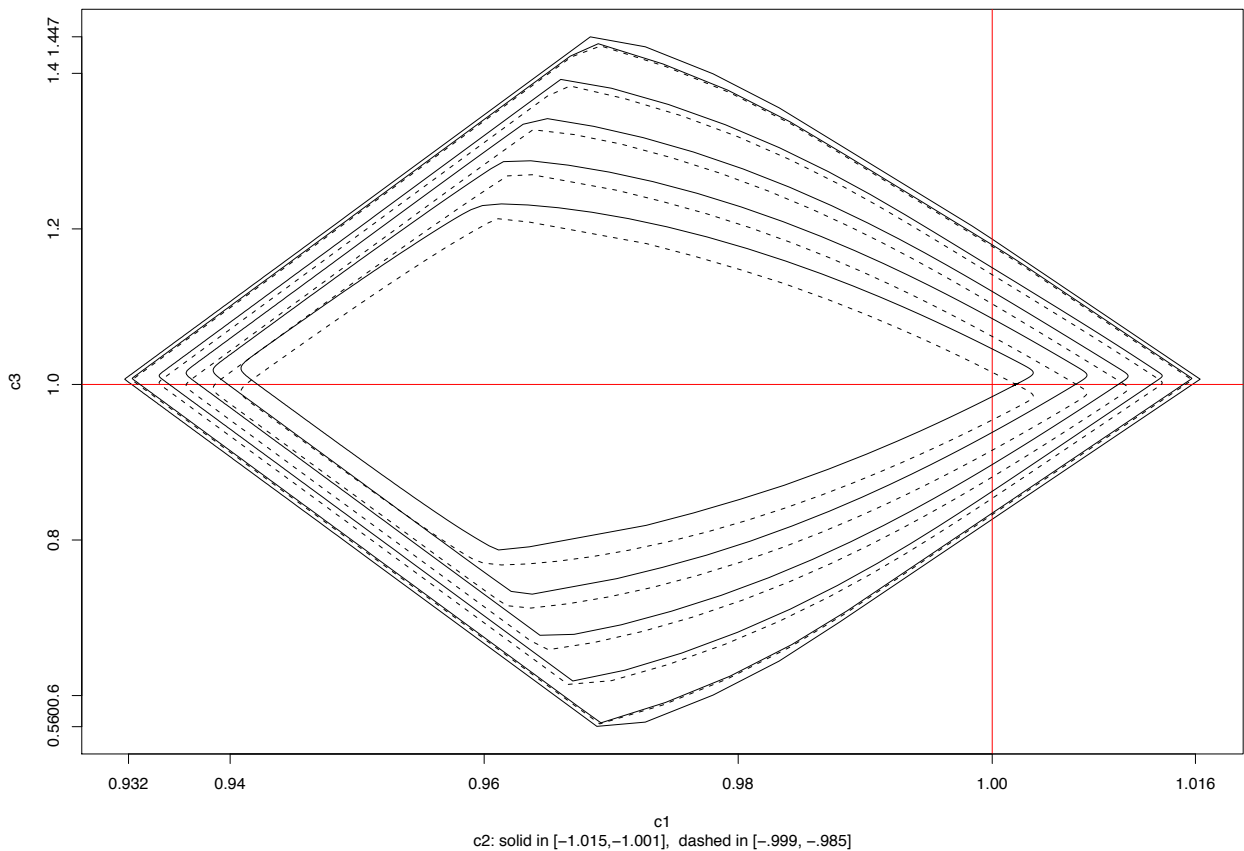


Figure 4: Normal case: Projections of the true set on  $(c_1, c_3)$ , range  $[-1.015, -0.985]$  of values of  $c_2$

Similarly, we can compute the exact inequalities corresponding to MMM for both the uniform and the normal case, and the details are given in Appendix B. As already mentioned in the introduction, 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 the values of  $v_0$  (all possible values in the uniform case, the restricted set mentioned above in the normal case) and powers of  $|x|$  for the normal case and of  $x/E(x)$  for the uniform case. We report sets for 500 powers for the uniform case and, for numerical reasons only 200 powers for the normal case, unusually large numbers of

<sup>5</sup>The discrepancies between the bounds on  $c_1$  and  $c_3$  in Figure 5 and the values reported in Remark 7 come from limitation imposed on  $x$  here.

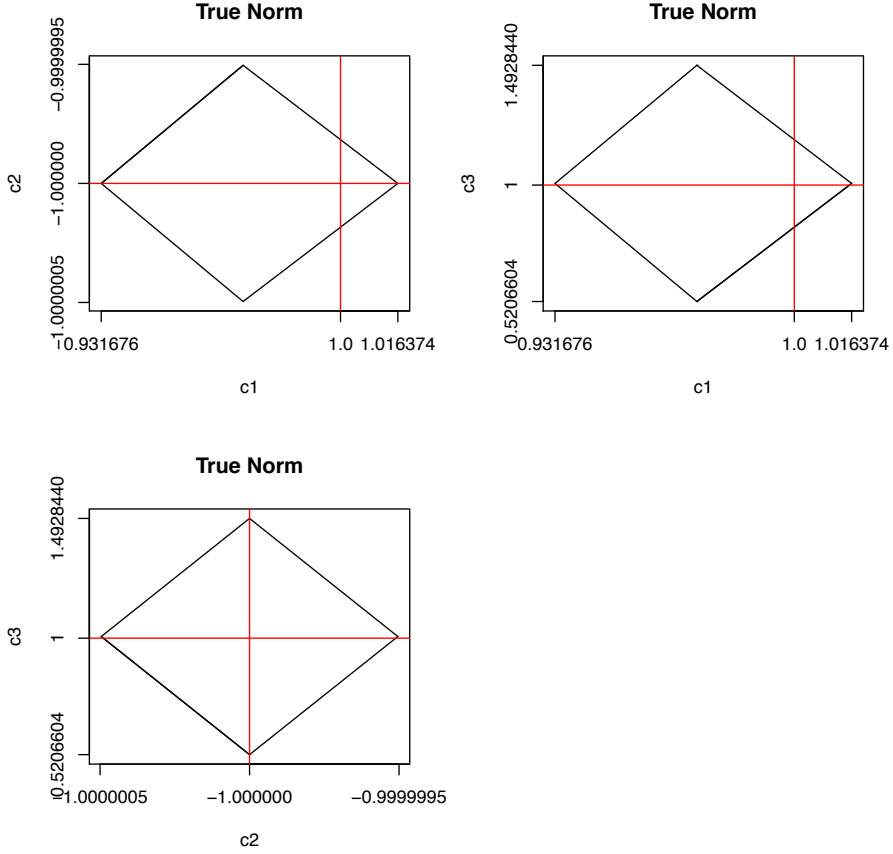


Figure 5: Normal case: Projections of the polyhedral approximation

moments, but recall that this is not an estimation problem. Figures 6 and 7 show the results. For the uniform case, there is a substantial improvement when moving from using two powers of  $x$  to 500, but 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  $C_w^{**}$  (.0356) is also more than three times the volume of  $C^*$  (.1119), as reported in Tables C.1 and C.2 (in Appendix C). For the normal case we also find a large improvement in moving from two to 200 powers of  $|x|$ , but more importantly,  $C_w^{**}$  is now very close to our numerically computed  $C^*$  and to its polyhedral approximation  $C_a^*$ , in particular if we consider the projections on the plane  $(c_1, c_3)$ , where  $C_w^{**}$  has exactly the same curved shape as  $C^*$  for values of  $c_2$  close to  $-1$  (See Figure 4).<sup>6</sup> This is remarkable, because obtaining  $C_w^{**}$  numerically does not require any approximation, contrary to obtaining  $C^*$  numerically.

<sup>6</sup>Note that this is still a polyhedron, only with many vertices.

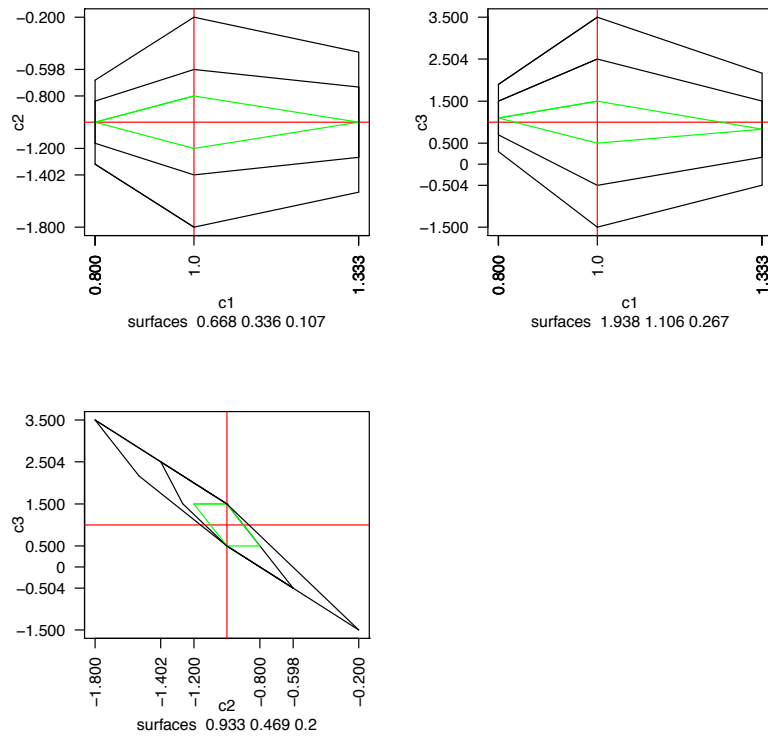


Figure 6: Uniform case: Projections of  $C_w^{**}$  with powers of 2 and 500

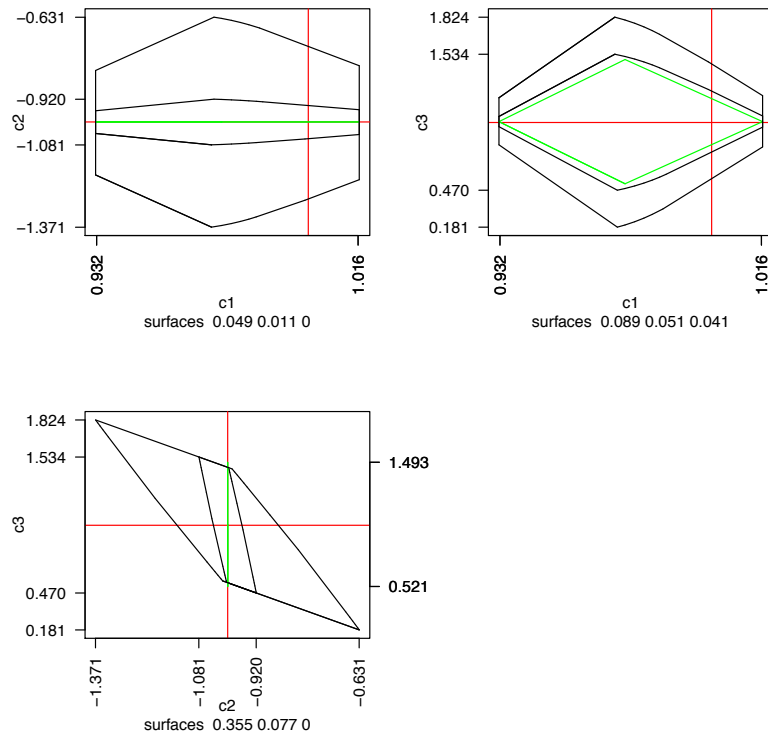


Figure 7: Normal case: Projections of  $C_w^{**}$  with powers of 2 and 200

## 4 Monte Carlo evidence

In this section, we replicate and extend the Monte Carlo design in MT with a twofold aim. First, we complete the analysis by presenting a larger range of sample sizes. We also 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  $\eta(x, v_0, v_1)$ . For the rest of the paper, we will somewhat loosely refer to the former as ‘nonparametric  $\eta$ ’ and to the latter as ‘exact  $\eta$ ’. Second, we compare the performances of MMD and MMM. The aim is to provide insights about the efficiency loss of using MMM, weighted against the reliance on an estimate of  $\eta$  in MMD.

Recall the linear model (LM):

$$y = \gamma_1 v + \gamma_2 x + \gamma_3 + \epsilon, \quad E(\epsilon|x, v) = 0, \quad (12)$$

where  $y$  is a continuous outcome,  $v$  an interval-measured continuous variable, and  $x$  a continuous regressor. We specify the parameter vector  $(\gamma_1, \gamma_2, \gamma_3) = (1, 1, -1)$  and  $(v, x, \epsilon)$  to be stochastically independent.

MT consider two cases. In the first,  $x$  and  $v$  are uniformly distributed. In the second, they are normally distributed. In both cases, they let  $\epsilon \sim N(0, 1)$ . In the uniform case, they set  $v \sim U[-2, 3]$  and  $x \sim U[0, 5]$ , and in the normal case,  $v \sim N(0, 2)$  and  $x \sim N(1, 4)$ . MT round  $v$  to obtain interval-measured  $v_0$  and  $v_1 = v_0 + 1$ . We report results for samples of sizes  $N = 100$ ,  $N = 200$ ,  $N = 800$ , and  $N = 20,000$ , as well as the true identified set. Recall that, in the normal case, what we call the true set really is a polyhedral approximation, whereas it is the sharp true set in the uniform case. In all cases, we employ 100 Monte Carlo replications. It would be desirable to consider a larger number of replications. However, here we follow MT and find that our results show sufficient stability to generate meaningful insights.

For MMD, we apply the simplified geometrical approach proposed in Section 3 to estimate the identified set with simulated data using sample analogues. In a first step, we need to estimate  $\eta$  nonparametrically. We use product kernels involving a normal kernel for the continuous variables. For practical reasons we treat  $v_0$  as an ordered factor in the uniform case, using a Wang and Van Ryzin kernel (see Hayfield and Racine, 2008), but as a continuous variable in the normal case. 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 variables ( $x$  in all cases,  $v_0$  in the normal case) by taking differences in the sample standard deviations into account. In the uniform case, 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 continuous variables 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.<sup>7</sup> The estimates with nonparametric  $\eta$  could be improved

---

<sup>7</sup>See Hayfield and Racine (2008) and the online documentation for the R function `npregbw` in the package `np`.

by performing LSCV for each MC replication.

For MMM, we use the  $2H$  inequalities from Equation (6) 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 subsequently increase the powers of scaled  $|x|$ , denoted by  $k$  in the function  $w$ . The powers we report are  $k = 2, 5, 20, 30$ . We further include a set of indicators for the values of  $v_0$ . We construct indicators for all possible values of  $v_0$  in the uniform case. In the normal case, we restrict the set of indicators to accommodate very low numbers of observations at the tails.

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 always 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>8</sup> Figures 8 to 13 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>9</sup> The green set is the true set, the rose sets are the estimates, and the solid black set 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.

#### 4.1 Findings: Modified Minimum Distance

Tables 1 and 2 report MT’s MMD results including their confidence regions,<sup>10</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 functional form of the exact  $\eta$  is in general unknown. In the uniform case, some characteristics of  $C^*$  are estimated very fast. This is the case for the number of vertices and the slopes of the sides, that is, the shapes of the projected sets. The polyhedral approximation performs well in the normal case. In most cases, the geometrical 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

---

<sup>8</sup>The quotation marks emphasize the fact that we merely summarize our results of the Monte Carlo studies. We leave the issue of inference aside in this paper.

<sup>9</sup>We adapt 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://www.business.uzh.ch/professorships/entrepreneurship/team/ullrich/mcappendix.pdf>

<sup>10</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.

precision using exact  $\eta$  over its nonparametric estimate for sample sizes of  $N = 20,000$  and for almost all sample sizes in the uniform case. In the uniform case, even with the smallest sample,  $N = 100$ , the mean estimated bounds using exact  $\eta$  are very close to the true bounds. The main benefit from increased sample sizes are 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. Table C.1 in Appendix C shows that, in the uniform case, 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  $N = 20,000$ .

Second, in the uniform case, column 2 in 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 8, and to a lesser extent Figure 10 with nonparametric  $\eta$ , demonstrates how well the MMD estimate picks up the shape of the true set even with very small sample sizes. In the normal case, this is true only for the projections on the  $v$  coefficient and the intercept.<sup>11</sup> MT does better in estimating the coefficient of  $x$  which is point-identified, whereas our estimates do not converge as closely to its true value even with the largest sample sizes.

Third, compared to MT, the geometrical 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. In the uniform case, the bounds reported by MT are too narrow and even their confidence intervals fail to cover the true bounds. The results are less clear cut in the normal case. We observe the largest differences between MT and the geometrical approach in the estimates for the intercept, in both the uniform and the normal cases, where MT's mean estimates and confidence intervals fail to cover the true bounds. Hence, MT is 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. For the uniform case, Table C.1 shows that the estimated and true polyhedra cover only about 17% of the volume of the Cartesian product of the corresponding intervals. 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 preliminary 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

---

<sup>11</sup>See the Online Appendix for the corresponding figures.

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

$$c_1 v_0 + c_2 x + c_3 \leq \Phi^{-1}[\eta(x, v_0, v_1)] \leq c_1 v_0 + c_2 x + c_3.$$

We have not (yet) worked out the exact functional form of  $\eta(x, v_0, v_1)$  for the BRM, and it might not be worth the trouble, since we have already documented the difference between exact and np  $\eta$  in the LM. In estimating  $\eta$  nonparametrically, we have experimented with simple nonparametric regression (ignoring the dichotomous character of  $y$ ), nonparametric conditional density estimation (R function `npcdens` in package `np`), 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. However, we have only obtained plausible results with `npreg` in the uniform case. We intend to experiment further, possibly using the local logit approach in Frölich (2006).

As do MT, overall we find wider confidence regions for the bounds in the BRM than in the LM. We make this observation even though the confidence regions are defined differently, since both our estimated bounds and the outer bounds of the respective confidence intervals are wider in the former. 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 lower bound of the  $x$  coefficient with smaller sample sizes. While not as apparent as in the LM, our results seem to converge faster to the true bounds in the BRM. Comparing the results for  $N = 800$ , the mean estimated bounds for the  $v$  coefficient are much closer using the geometrical 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 except the lower bound of the  $x$  coefficient. 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.

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

Coefficient	Lower/Upper bound	Linear Model		Binary Response Model		
		MT	true $\eta$	np $\eta$	MT	np $\eta$
<i>True: .800</i>						
$v$	N = 100	.98 ( <i>.87</i> )	.795 [.688, .895]	.822 [.715, .947]	1.05 (.73)	.778 [.563, 1.077]
	N = 200	.99 ( <i>.92</i> )	.799 [.743, .863]	.834 [.738, .949]	.95 ( <i>.81</i> )	.816 [.602, 1.060]
	N = 800	.97	.802 [.771, .839]	.812 [.752, .871]	.85	.801 [.683, .954]
	N = 20000		.800 [.794, .807]	.813 [.768, .852]		.851 [.778, .907]
<i>True: 1.333</i>						
	N = 100	1.11 ( <i>1.23</i> )	1.325 [1.147, 1.491]	1.139 [.935, 1.310]	1.34 (1.91)	1.155 [.829, 1.451]
	N = 200	1.068 ( <i>1.17</i> )	1.332 [1.239, 1.438]	1.179 [.999, 1.352]	1.07 ( <i>1.29</i> )	1.171 [.806, 1.410]
	N = 800	1.27	1.336 [1.285, 1.399]	1.243 [1.150, 1.334]	1.14	1.273 [.957, 1.525]
	N = 20000		1.333 [1.323, 1.344]	1.299 [1.224, 1.366]		1.391 [1.129, 1.553]
<i>True: -1.200</i>						
$x$	N = 100	-1.03 ( <i>-1.14</i> )	-1.192 [-1.312, -1.079]	-1.135 [-1.329, -.933]	-1.21 (-1.77)	-868 [-1.092, -.597]
	N = 200	-1.01 ( <i>-1.08</i> )	-1.211 [-1.284, -1.137]	-1.143 [-1.265, -.975]	-1.01 (-1.18)	-907 [-1.096, -.704]
	N = 800	-1.14	-1.197 [-1.234, -1.152]	-1.172 [-1.246, -1.097]	-1.09	-1.044 [-1.180, -.825]
	N = 20000		-1.199 [-1.210, -1.189]	-1.186 [-1.226, -1.154]		-1.183 [-1.285, -1.018]
<i>True: -.800</i>						
	N = 100	-0.93 ( <i>-.82</i> )	-0.787 [-0.920, -.659]	-0.865 [-1.064, -.661]	-0.96 (-.68)	-524 [-.761, -.247]
	N = 200	-0.84 ( <i>-.87</i> )	-0.807 [-.889, -.736]	-0.854 [-.998, -.740]	-0.90 (-.75)	-586 [-.796, -.399]
	N = 800	-0.84	-0.795 [-.845, -.749]	-0.831 [-.914, -.769]	-0.84	-649 [-.779, -.531]
	N = 20000		-0.799 [-.810, -.790]	-0.813 [-.850, -.775]		-751 [-.818, -.672]
<i>True: .500</i>						
Intercept	N = 100	.79 (.48)	.474 [.061, .827]	.610 [.073, 1.093]	.83 (.35)	-256 [-1.176, .594]
	N = 200	.84 (.61)	.522 [.301, .752]	.541 [.134, 1.021]	.80 (.53)	-114 [-.990, .624]
	N = 800	.58	.484 [.372, .617]	.531 [.329, .747]	.68	.052 [-.479, .386]
	N = 20000		.500 [.475, .527]	.510 [.378, .650]		.256 [-.064, .507]
<i>True: 1.500</i>						
	N = 100	1.08 ( <i>1.38</i> )	1.487 [1.168, 1.835]	1.491 [1.004, 1.976]	1.22 (2.09)	.941 [-.226, 1.735]
	N = 200	1.01 ( <i>1.22</i> )	1.533 [1.322, 1.739]	1.450 [1.042, 1.841]	.94 ( <i>1.35</i> )	.944 [.016, 1.717]
	N = 800	1.31	1.490 [1.382, 1.605]	1.497 [1.299, 1.698]	1.24	1.316 [.548, 1.721]
	N = 20000		1.499 [1.476, 1.528]	1.506 [1.364, 1.640]		1.523 [1.087, 1.801]

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.



Table 2: Monte Carlo Analysis — MMD, Normal Case — One-dimensional Bounds

Coefficient	Lower/Upper bound	MT	true $\eta$	np $\eta$
$v$	<i>True: .932, Approx.: .932</i>			
	N = 100		.848 [ .747, .941 ]	.856 [ .749, .961 ]
	N = 200	.97 (.88)	.855 [ .780, .936 ]	.868 [ .789, .942 ]
	N = 800	.96 (.91)	.867 [ .831, .912 ]	.880 [ .837, .927 ]
	N = 20000	.93	.885 [ .877, .895 ]	.895 [ .867, .926 ]
	<i>True: 1.016, Approx.: 1.016</i>			
	N = 100		1.126 [ .983, 1.248 ]	1.115 [ .983, 1.253 ]
	N = 200	1.07 (1.19)	1.112 [ 1.019, 1.215 ]	1.093 [ .994, 1.209 ]
	N = 800	1.02 (1.09)	1.088 [ 1.048, 1.140 ]	1.068 [ 1.015, 1.119 ]
	N = 20000	1.08	1.057 [ 1.043, 1.068 ]	1.039 [ .990, 1.077 ]
$x$	<i>True: -1.015, Approx.: -1</i>			
	N = 100		-1.092 [ -1.175, -1.021 ]	-1.087 [ -1.174, -1.016 ]
	N = 200	-1.03 (-1.09)	-1.083 [ -1.146, -1.017 ]	-1.074 [ -1.145, -1.010 ]
	N = 800	-1.02 (-1.04)	-1.078 [ -1.111, -1.047 ]	-1.068 [ -1.099, -1.031 ]
	N = 20000	-1.00	-1.061 [ -1.069, -1.053 ]	-1.053 [ -1.078, -1.029 ]
	<i>True: -.985, Approx.: -1</i>			
	N = 100		-.895 [ -.977, -.819 ]	-.901 [ -.982, -.819 ]
	N = 200	-.95 (-.88)	-.906 [ -.967, -.843 ]	-.916 [ -.975, -.851 ]
	N = 800	-.99 (-.94)	-.924 [ -.956, -.893 ]	-.934 [ -.968, -.904 ]
	N = 20000	-.99	-.940 [ -.948, -.933 ]	-.951 [ -.975, -.931 ]
Intercept	<i>True: .560, Approx.: .521</i>			
	N = 100		.518 [ .319, .696 ]	.540 [ .342, .750 ]
	N = 200	.90 (.73)	.519 [ .386, .665 ]	.581 [ .364, .756 ]
	N = 800	.91 (.80)	.519 [ .447, .585 ]	.576 [ .405, .749 ]
	N = 20000	.93	.519 [ .508, .532 ]	.583 [ .368, .768 ]
	<i>True: 1.447, Approx.: 1.493</i>			
	N = 100		1.485 [ 1.293, 1.688 ]	1.460 [ 1.217, 1.730 ]
	N = 200	1.08 (1.25)	1.486 [ 1.367, 1.627 ]	1.455 [ 1.244, 1.671 ]
	N = 800	1.03 (1.14)	1.484 [ 1.412, 1.549 ]	1.427 [ 1.268, 1.581 ]
	N = 20000	1.04	1.483 [ 1.468, 1.498 ]	1.404 [ 1.192, 1.568 ]

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. Intervals in italic indicate the failure to cover the true value of the corresponding bound. We include both the numerically obtained true set and the polyhedral approximation. The former is plotted in Figure 4. The MC study with the largest sample completed within 72 seconds, using nonparametric  $\eta$  with bandwidth computed for only one replication for  $N = 200$  and scaled down for  $N = 800$  and  $N = 20,000$ .

## 4.2 Findings: Modified Method of Moments

Tables 3 and 4 report our MMM results for alternative values of  $k$ , the highest powers of scaled  $x$  used in  $w$ . We consider  $k = 2, 5, 20, 30$ . In general, we find that the convergence of the estimated bounds to the true bounds is improved as the sample size increases. As in the MMD case, we find that a large sample size of  $N = 20,000$  is not always enough to achieve convergence to the true bounds. These results are more clear in the uniform case. Moreover, we also find that the gains obtained from increasing  $k$  are limited. That is, there is a noticeable improvement in the convergence to the true bounds (as  $N$  increases) from adjusting  $k$  from 2 to 20 but this improvement

is more modest for  $k > 20$ . This is reassuring given the considerable computational costs attached to the estimation with large  $k$ . Finally, we find that MMD outperforms MMM, even considering the cost of estimating  $\eta$  nonparametrically.

More specifically, column 2 in Table 3 shows, for the uniform case, the estimation of the bounds setting  $k = 2$ . With  $N = 20,000$ , the estimated bounds are very close to their true values (i.e. the first of the two true values reported) and very precisely estimated. Note that in general, for smaller sample sizes, the estimated bounds do not include the true bounds, a phenomenon that becomes clear as  $k$  increases. Only for the estimated bounds for the coefficient of  $v$ , a large sample size guarantees the convergence to the true bounds. Moreover, as highlighted above, the results tend to vary less as  $k$  is increased beyond 20. The results for the normal case are less clear cut.

Table 3: Monte Carlo Analysis — MMM, Uniform Case — 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	[-1.888, -1.531]	-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	[-1.608, -1.439]	-1.412	[-1.495, -1.319]	-1.398	[-1.494, -1.292]
	N = 20000	-1.787	[-1.804, -1.767]	-1.552	[-1.565, -1.536]	-1.434	[-1.450, -1.419]	-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	[-.537, -.403]	-.580	[-.664, -.491]	-.594	[-.696, -.495]
	N = 20000	-.213	[-.236, -.193]	-.447	[-.466, -.432]	-.566	[-.583, -.549]	-.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	[-1.041, -.537]	-.497	[-.775, -.231]	-.463	[-.765, -.202]
	N = 20000	-1.450	[-1.516, -1.377]	-.866	[-.923, -.807]	-.572	[-.629, -.511]	-.539	[-.597, -.472]
	<i>True: 3.500, 2.504</i>								
	N = 100	2.775	[2.147, 3.438]	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, 3.012]	2.350	[1.871, 2.832]	2.306	[1.836, 2.822]
	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	[2.815, 2.907]	2.570	[2.521, 2.624]	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  $k = 500$ .

Table 4: Monte Carlo Analysis — MMM, Normal Case — One-dimensional Bounds

Coefficient	Lower/Upper bound	$k = 2$		$k = 5$		$k = 20$		$k = 30$		
$v$	<i>True: .932, .932</i>									
	N = 100	.830	[.704, .951]	.841	[.724, .958]	.847	[.724, .970]	.850	[.724, .999]	
	N = 200	.814	[.747, .899]	.820	[.752, .909]	.830	[.759, .916]	.828	[.735, .924]	
	N = 800	.807	[.770, .851]	.812	[.774, .856]	.820	[.779, .879]	.827	[.779, .903]	
	N = 20000	.811	[.803, .820]	.812	[.803, .821]	.813	[.804, .823]	.819	[.804, .863]	
	<i>True: 1.016, 1.016</i>									
	N = 100	1.197	[.959, 1.392]	1.179	[.980, 1.369]	1.165	[.967, 1.351]	1.160	[.966, 1.349]	
	N = 200	1.219	[1.095, 1.346]	1.202	[1.089, 1.325]	1.192	[1.079, 1.310]	1.185	[1.041, 1.340]	
	N = 800	1.204	[1.136, 1.285]	1.192	[1.127, 1.273]	1.188	[1.125, 1.258]	1.187	[1.124, 1.255]	
	N = 20000	1.179	[1.163, 1.196]	1.177	[1.162, 1.193]	1.176	[1.161, 1.189]	1.175	[1.161, 1.189]	
	$x$	<i>True: -1.371, -1.081</i>								
		N = 100	-1.449	[-1.905, -1.146]	-1.254	[-1.518, -.999]	-1.149	[-1.397, -.891]	-1.138	[-1.397, -.839]
		N = 200	-1.492	[-1.753, -1.295]	-1.268	[-1.473, -1.096]	-1.159	[-1.417, -.920]	-1.136	[-1.353, -.839]
		N = 800	-1.539	[-1.663, -1.430]	-1.272	[-1.393, -1.174]	-1.147	[-1.343, -.928]	-1.130	[-1.318, -.876]
N = 20000		-1.586	[-1.613, -1.560]	-1.280	[-1.301, -1.262]	-1.128	[-1.212, -1.023]	-1.114	[-1.210, -.982]	
<i>True: -.631, -.920</i>										
N = 100		-.580	[-.969, -.178]	-.747	[-.972, -.467]	-.837	[-1.154, -.531]	-.848	[-1.122, -.560]	
N = 200		-.480	[-.660, -.308]	-.720	[-.876, -.553]	-.834	[-1.062, -.634]	-.835	[-1.084, -.535]	
N = 800		-.461	[-.567, -.342]	-.729	[-.813, -.621]	-.850	[-1.063, -.693]	-.863	[-1.109, -.704]	
N = 20000		-.413	[-.441, -.386]	-.721	[-.742, -.697]	-.874	[-.968, -.783]	-.893	[-1.030, -.788]	
Intercept		<i>True: .181, .470</i>								
		N = 100	.414	[-.080, 1.017]	.538	[.214, .888]	.619	[.256, 1.037]	.633	[.274, 1.138]
		N = 200	.212	[-.126, .583]	.410	[.120, .675]	.533	[.251, .843]	.521	[.222, .977]
		N = 800	.096	[-.057, .263]	.340	[.218, .455]	.465	[.294, .751]	.490	[.293, .817]
	N = 20000	-.042	[-.075, .003]	.260	[.227, .289]	.414	[.312, .501]	.446	[.321, .679]	
	<i>True: 1.824, 1.534</i>									
	N = 100	1.655	[1.189, 2.113]	1.504	[1.132, 1.903]	1.404	[1.026, 1.767]	1.387	[.963, 1.766]	
	N = 200	1.757	[1.446, 2.108]	1.575	[1.285, 1.864]	1.460	[1.172, 1.711]	1.447	[1.043, 1.761]	
	N = 800	1.914	[1.748, 2.064]	1.668	[1.531, 1.840]	1.534	[1.265, 1.750]	1.504	[1.015, 1.719]	
	N = 20000	2.041	[1.998, 2.080]	1.742	[1.712, 1.772]	1.591	[1.481, 1.672]	1.566	[1.426, 1.670]	

Notes: We report two sets of ‘confidence intervals’, one for the lower bound and one for the upper bound of the estimated interval. Intervals in italic indicate the failure to cover the true value of the corresponding bound. The first set of true bounds corresponds to the true set with  $k = 2$  and the second to  $k = 200$ .

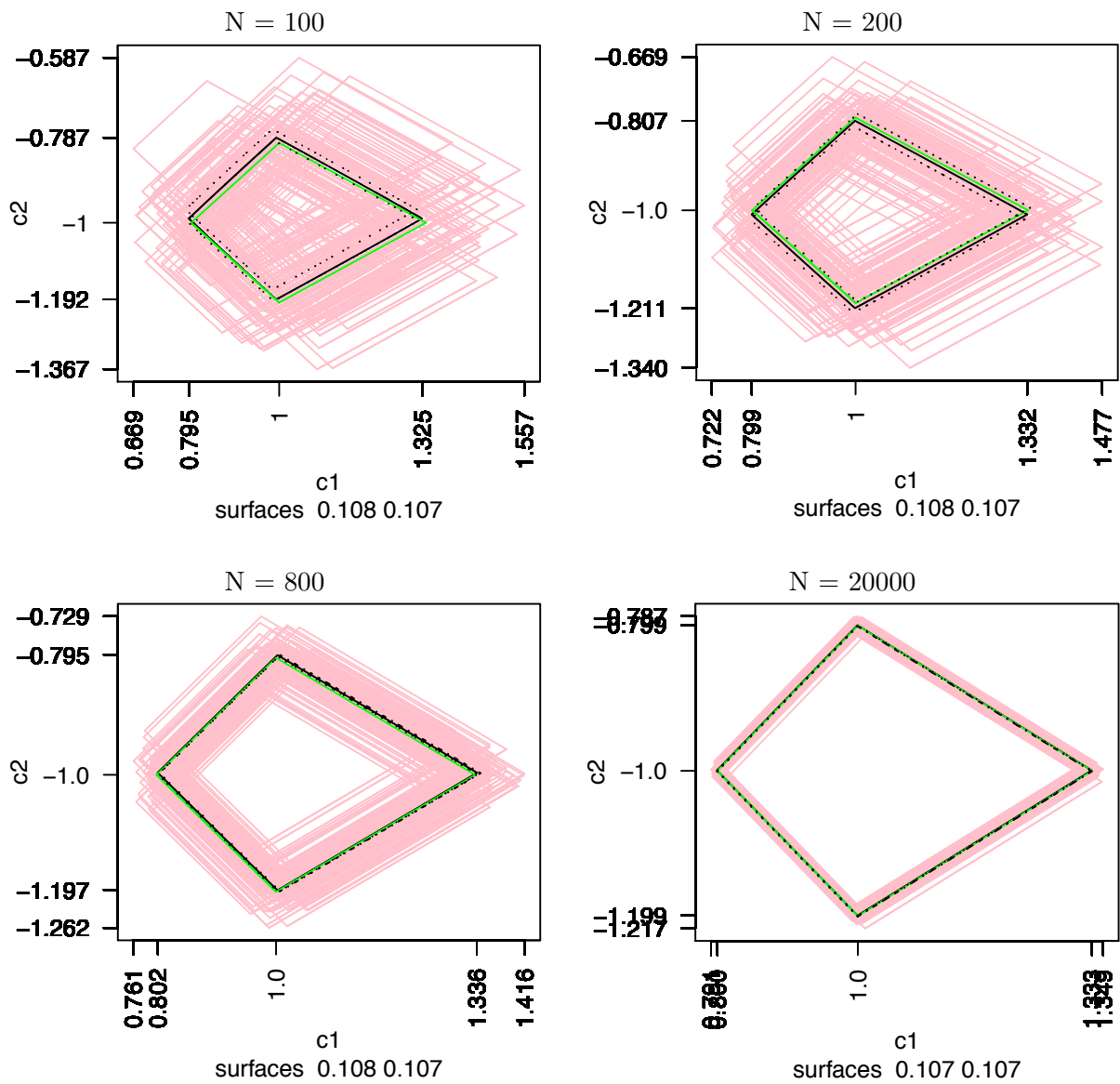


Figure 8: MMD, uniform case — True  $\eta$  — Projections on  $v$  and  $x$

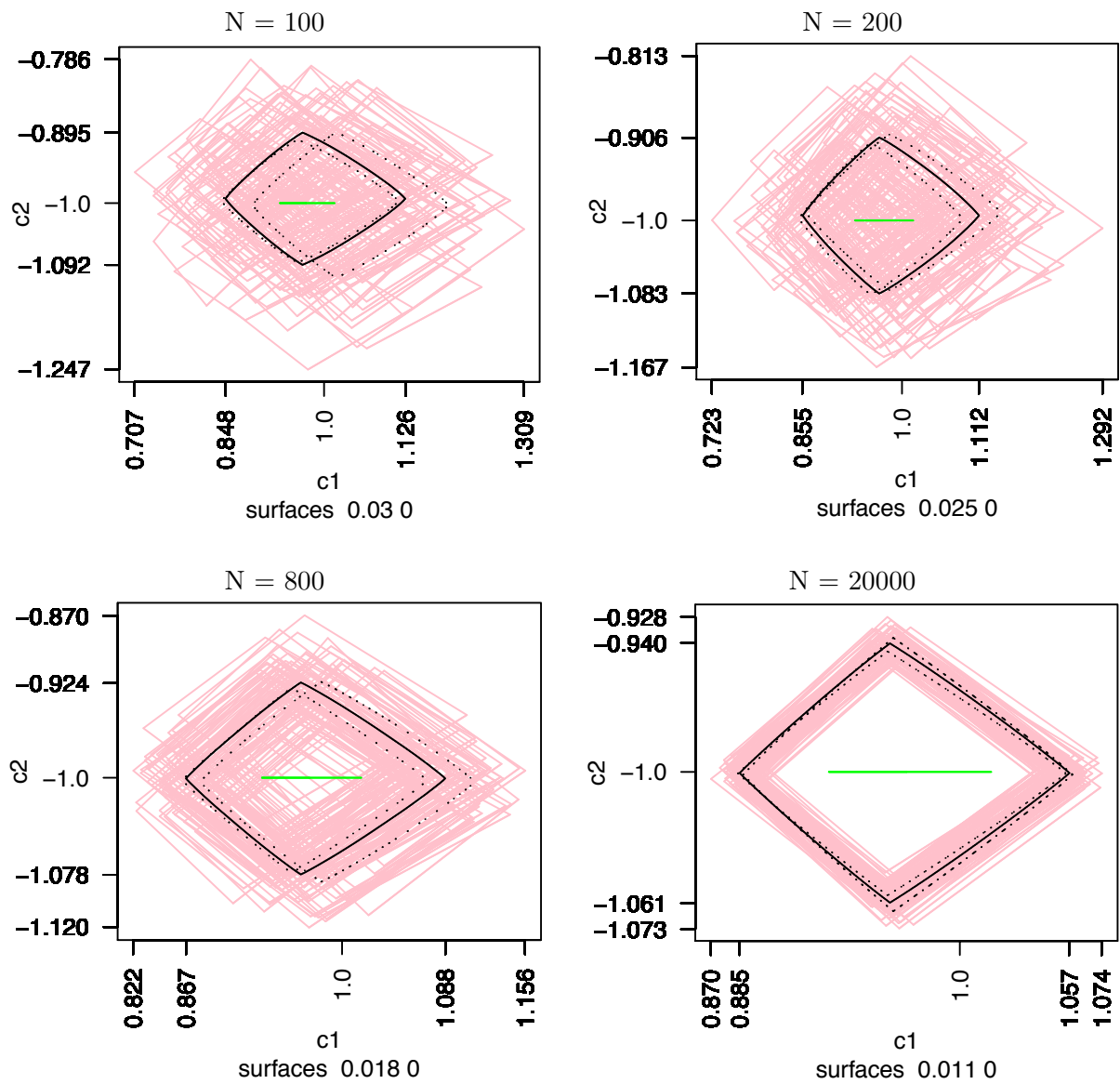


Figure 9: MMD, normal case — True  $\eta$  — Projections on  $v$  and  $x$

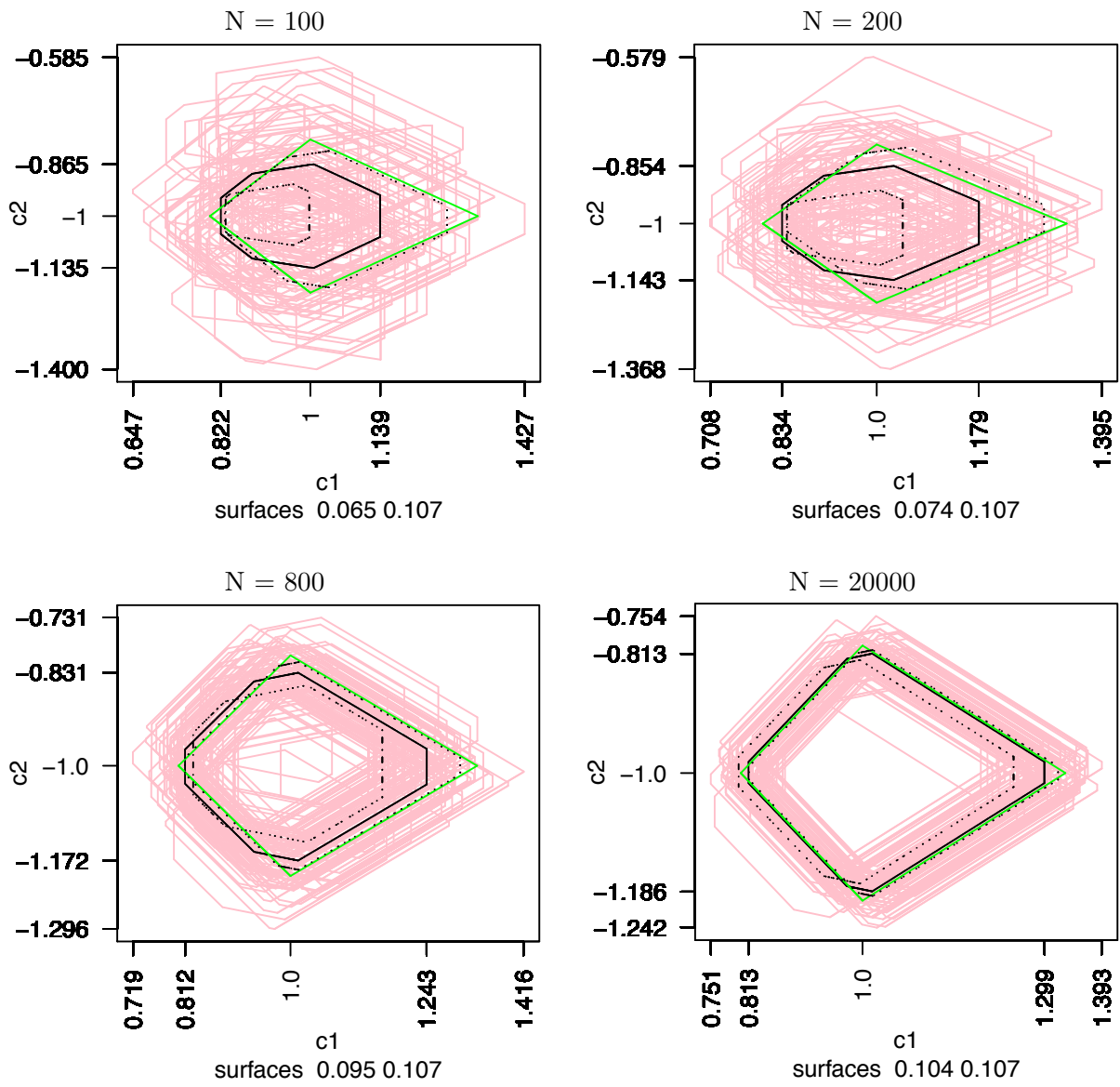


Figure 10: MMD, uniform case — Nonparametric  $\eta$  — Projections on  $v$  and  $x$

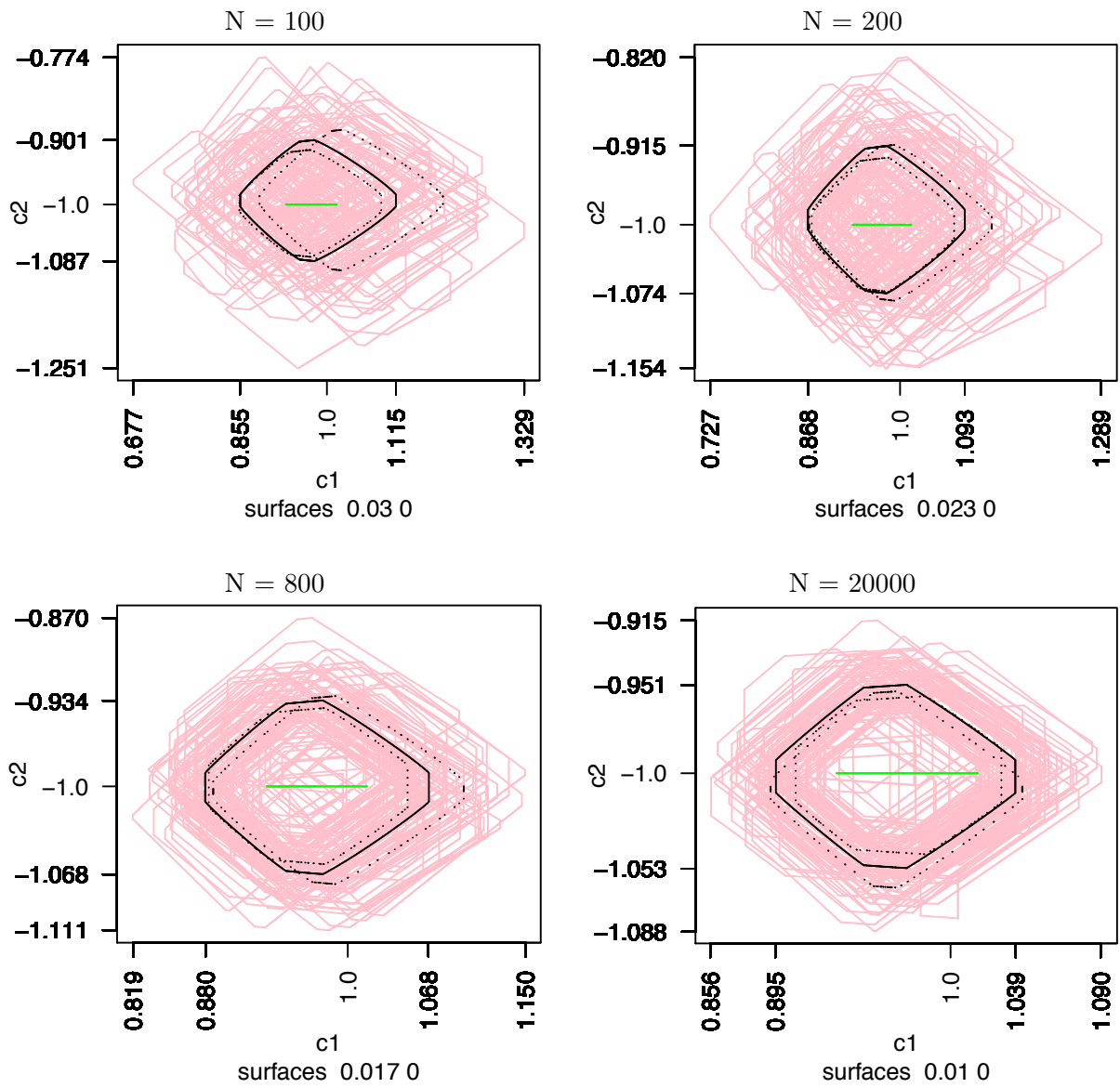


Figure 11: MMD, normal case — Nonparametric  $\eta$  — Projections on  $v$  and  $x$

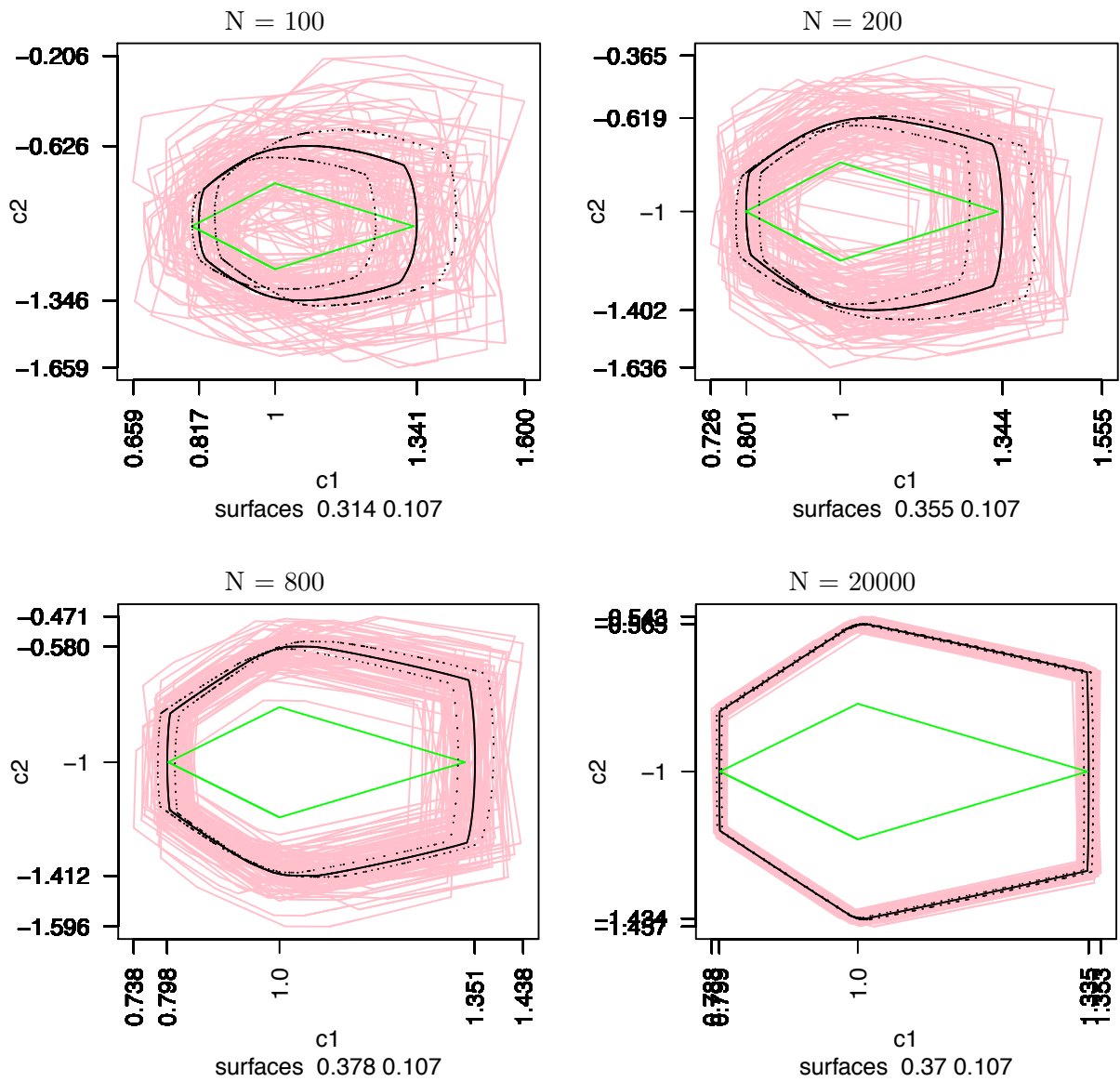


Figure 12: MMM, uniform case —  $k = 20$  — Projections on  $v$  and  $x$



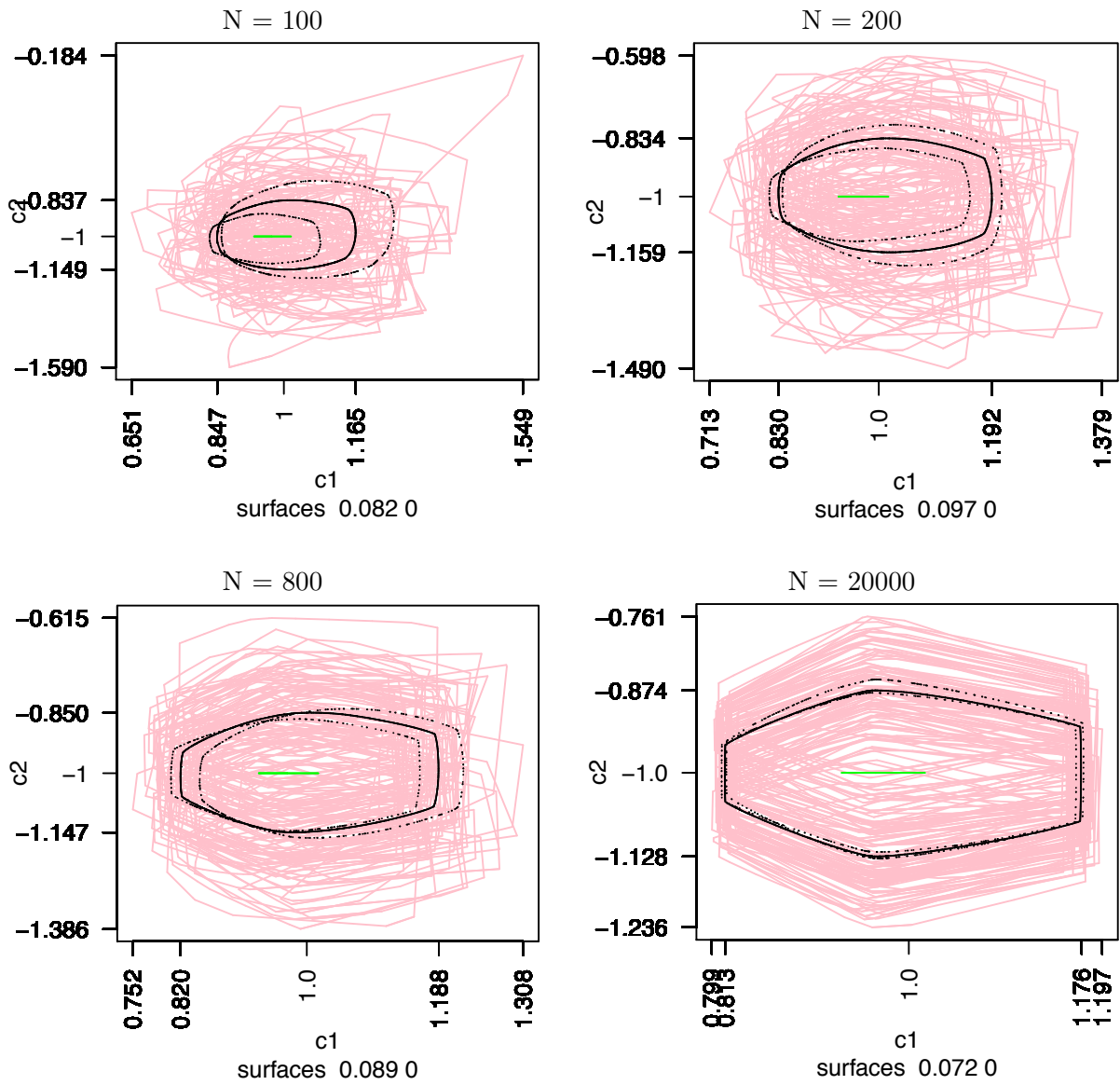


Figure 13: MMM, normal case —  $k = 20$  — Projections on  $v$  and  $x$

## Appendix A Asymptotic MMD criterion for the normal case

We want to compute the function

$$Q(c, \eta) = Q_1(c, \eta) + Q_0(c, \eta)$$

with

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

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

and

$$f(x, v, c) = c_1v + c_2x + c_3$$

$$x \sim N(1, 4), \quad v \sim N(0, 2)$$

$$\eta(x, v_0, v_1) = 1 - x + E(v|v_0) = 1 - x + \bar{v}_0 = 1 - x + \sqrt{2} \frac{\phi(v_0/\sqrt{2}) - \phi(v_1/\sqrt{2})}{\Phi(v_1/\sqrt{2}) - \Phi(v_0/\sqrt{2})}$$

and hence  $v_0$  takes all positive and negative integer values with probabilities

$$P[v_0 = n] = P[v_1 = n + 1] = P[n < v < n + 1] = \Phi\left[(n + 1)/\sqrt{2}\right] - \Phi\left[n/\sqrt{2}\right]$$

The first 12 values of  $P[v_0 = n]$  (for  $n = 0, \dots, 11$ ) are:

0	1	2	3	4	5	6	7	8	9	10	11
.260	.161	.0617	.0146	$2.14e-3$	$1.92e-4$	$1.07e-5$	$3.64e-7$	$7.61e-9$	$9.75e-11$	$7.65e-13$	$3.66e-15$

Thus we will compute

$$Q_1(c, \eta) = \sum_{v_1=-11}^{12} P[v_0 = v_1 - 1] E_x\{\mathbf{1}[f(x, v_1, c) < \eta(x, v_0, v_1)]\}$$

$$Q_0(c, \eta) = \sum_{v_0=-12}^{11} P[v_1 = v_0 + 1] E_x\{\mathbf{1}[f(x, v_0, c) > \eta(x, v_0, v_1)]\}.$$

Condition  $f(x, v_1, c) < \eta(x, v_0, v_1)$  is equivalent to

$$c_1v_1 + c_2x + c_3 < 1 - x + \bar{v}_0$$

or, with  $a = c_2 + 1$ ,

$$ax < d_1(c, v_1) := 1 + \bar{v}_0 - c_1v_1 - c_3. \tag{13}$$

Condition  $f(x, v_0, c) > \eta(x, v_0, v_1)$  is equivalent to

$$c_1 v_0 + c_2 x + c_3 > 1 - x + \bar{v}_0$$

or

$$ax > d_0(c, v_0) := 1 + \bar{v}_0 - c_1 v_0 - c_3 \quad (14)$$

1. Case  $a > 0$

Then, given  $v_1$  and condition (13),  $x$  varies between  $-\infty$  and  $\delta_1 := d_1(c, v_1)$ . Thus,

$$Q_1^v(c, v_1) = \Phi\left(\frac{\delta_1 - E(x)}{\sqrt{V(x)}}\right) = \Phi\left(\frac{\delta_1 - 1}{2}\right).$$

Given  $v_0$  and condition (14),  $x$  varies between  $\delta_0 := d_0(c, v_0)/a$  and  $+\infty$ . Thus,

$$Q_0^v(c, v_0) = 1 - \Phi\left(\frac{\delta_0 - 1}{2}\right).$$

2. Case  $a < 0$

Then, given  $v_1$  and condition (13),  $x$  varies between  $\delta_1 := d_1(c, v_1)/a$  and  $+\infty$ , and

$$Q_1^v(c, v_1) = 1 - \Phi\left(\frac{\delta_1 - 1}{2}\right).$$

Given  $v_0$  and condition (14),  $x$  varies between  $-\infty$  and  $\delta_0 := d_0(c, v_0)/a$  and,

$$Q_0^v(c, v_0) = \Phi\left(\frac{\delta_0 - 1}{2}\right).$$

3. Case  $a = 0$

Then condition (13) imposes no restriction on  $x$  and two subcases must be distinguished.

(a)  $d_1(c, v_1) \leq 0$ . Then  $Q_1^v(c, v_1) = 0$ .

(b)  $d_1(c, v_1) > 0$ . Then  $Q_1^v(c, v_1) = 1$ .

Condition (14) imposes no restriction on  $x$  either and two subcases must be distinguished.

(c)  $d_0(c, v_0) \geq 0$ . Then  $Q_0^v(c, v_0) = 0$ .

(d)  $d_0(c, v_0) < 0$ . Then  $Q_0^v(c, v_0) = 1$ .

**Computational details** The bisection method used to produce Figure 4 can be described as follows. Given a value of  $c_2$ , we first generate a grid of  $(c_1, c_3)$  values, check which of the resulting  $(c_1, c_2, c_3)$  belong to  $C^*$ , that is, which satisfy  $Q(c, \eta) = 0$ , and take the average of these values as reference point. Then, for 800 angles from 0 to  $2\pi$  we obtain the boundary point in the corresponding direction starting from the reference point and a step length equal to its norm. We use a minimum relative step size of  $10^{-5}$  and a threshold value of  $10^{-18}$  for the function value.

Of course, for estimation this method would be prohibitively time-consuming. The polyhedral approximation uses a tolerance of  $1.5^{-8}$  in deciding whether an inequality is satisfied or not (this plays the role of  $\epsilon_N$  in equation (5) and it is needed even to compute the true set, or better said, a numerical approximation thereof). A further tolerance specification which is needed is the tolerance for computing the rank of a matrix (R function rankMatrix) when looking for intersections of planes in  $\mathbb{R}^3$ . This was set to  $10^{-6}$ . All these choices were made by trial and error and it is difficult for us to assess their impact on the results.

## Appendix B True MMM inequalities

### Appendix B.1 Uniform case

We want to compute the exact form of the following inequalities

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

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

where

$$f(x, v, c) = c_1v + c_2x + c_3,$$

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

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.

$$x \sim U[0, 5], \quad v \sim U[-2, 3].$$

Hence,  $v_0$  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>12</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 need to complete our task 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.  $E(v) = 5/2 - 2 = 1/2$ .

---

<sup>12</sup>Normalizations can be useful to avoid numerical problems: for instance use powers of  $x/E(x)$  or of  $x/\sigma(x)$ , with  $\sigma^2(x) = V(x)$ .

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

## Appendix B.2 Normal case

We want to compute the exact form of the following inequalities

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

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

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.

$$x \sim N(1, 4), \quad v \sim N(0, 2)$$

and hence  $v_0$  takes all positive and negative integer values with probabilities

$$P[v_0 = n] = P[v_1 = n + 1] = P[n < v < n + 1] = \Phi\left[(n + 1)/\sqrt{2}\right] - \Phi\left[n/\sqrt{2}\right]$$

The first 12 values of  $P[v_0 = n]$  (for  $n = 0, \dots, 11$ ) are listed in Appendix 1.

Now we have

$$E(v|v_0) = \sqrt{2} \frac{\phi(v_0/\sqrt{2}) - \phi(v_1/\sqrt{2})}{\Phi(v_1/\sqrt{2}) - \Phi(v_0/\sqrt{2})}.$$

The "instruments"  $w$  are defined as  $w' = (w'_v, w'_x)$ , with  $w_v$  a complete set of indicators for the values -12 to +11 of  $v_0$ , and  $w_{Kx}$  a vector of the first  $K$  powers of  $|x|$ :  $w'_x = (|x|, |x|^2, \dots, |x|^K)$ .<sup>13</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 need to complete our task 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.  $E(v) = 0$ .
2.  $E(x) = 1$ .
3.  $E(v_0) = -1/2$ , since  $v_0 + 1/2$  takes pairs of opposite values  $\{-n + 1/2, n - 1/2\}$  (e.g. -12+1/2, 11+1/2) with equal probabilities and thus has mean 0.
4.  $E(w_v)$ : this is a vector with 24 components  $E(1[v_0 = n]) = P[v_0 = n] = \Phi[(n + 1)/\sqrt{2}] - \Phi[n/\sqrt{2}]$ .
5.  $E(w_x)$ : this is a vector with  $K$  components  $E(|x|^k)$ ,  $k = 1, \dots, K$ . There is a formula for this, but it is messy, and we might use MC approximation.

---

<sup>13</sup>Here we cannot take powers of  $x$  itself, because  $w$  must be a vector of positive functions. Normalizations can be useful to avoid numerical problems: for instance use powers of  $x/E(x)$  or of  $x/\sigma(x)$ , with  $\sigma^2(x) = V(x)$ .

6.  $E(w_v v)$  : this is a vector with 24 components

$$\begin{aligned} E(v1[v_0 = n]) &= E(E(v|v_0 = n)1[v_0 = n]), \quad n = -12, \dots, 11, \\ &= E\left(\sqrt{2} \frac{\phi(v_0/\sqrt{2}) - \phi(v_1/\sqrt{2})}{\Phi(v_1/\sqrt{2}) - \Phi(v_0/\sqrt{2})} 1[v_0 = n]\right), \end{aligned}$$

since  $E(v|v_0) = \sqrt{2} \frac{\phi(v_0/\sqrt{2}) - \phi(v_1/\sqrt{2})}{\Phi(v_1/\sqrt{2}) - \Phi(v_0/\sqrt{2})}$ . Thus

$$\begin{aligned} E(v1[v_0 = n]) &= \sqrt{2} \frac{\phi(n/\sqrt{2}) - \phi[(n+1)/\sqrt{2}]}{\Phi[(n+1)/\sqrt{2}] - \Phi[n/\sqrt{2}]} \left(\Phi[(n+1)/\sqrt{2}] - \Phi[n/\sqrt{2}]\right) \\ &= \sqrt{2} \left\{ \phi(n/\sqrt{2}) - \phi[(n+1)/\sqrt{2}] \right\}. \end{aligned}$$

7.  $E(w_v v_0)$  : this is a vector with 24 components

$$\begin{aligned} E(v_0 1[v_0 = n]) &= nP[v_0 = n], \quad n = -12, \dots, 11, \\ &= n \left( \Phi[(n+1)/\sqrt{2}] - \Phi[n/\sqrt{2}] \right), \end{aligned}$$

and  $E(w_v v_1) = E(w_v (v_0 + 1))$  has components  $(n+1) \left( \Phi[(n+1)/\sqrt{2}] - \Phi[n/\sqrt{2}] \right)$ .

8.  $E(w_K x)$  : this is a vector with  $K$  components  $E(|x|^k x)$ ,  $k = 1, \dots, K$ . If  $k$  is even this is  $E(x^{k+1})$ , since then  $|x|^k = x^k$ ; if  $k$  is odd this is  $E(x^{k+1} \text{sgn}(x))$ , since then  $|x|^k = x^{k-1}|x|$  and  $|x|^k x = x^k|x| = x^{k+1} \text{sgn}(x)$ . Unfortunately, this seems rather complicated ... but we may want to give it a try (at worst we can approximate it by MC, in the same loop as  $E(|x|^k)$ ).

Finally, a choice of  $w_x$  that greatly simplifies computations 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

$$\begin{aligned} E(w_{xk} x) &= E[w_{xk} E(x|x_{k-1} < x \leq x_k)] \\ &= E\left[w_{xk} \left(1 + 2 \frac{\phi\left(\frac{x_{k-1}-1}{2}\right) - \phi\left(\frac{x_k-1}{2}\right)}{\Phi\left(\frac{x_k-1}{2}\right) - \phi\left(\frac{x_{k-1}-1}{2}\right)}\right)\right] \\ &= \left[\Phi\left(\frac{x_k-1}{2}\right) - \phi\left(\frac{x_{k-1}-1}{2}\right)\right] \left(1 + 2 \frac{\phi\left(\frac{x_{k-1}-1}{2}\right) - \phi\left(\frac{x_k-1}{2}\right)}{\Phi\left(\frac{x_k-1}{2}\right) - \phi\left(\frac{x_{k-1}-1}{2}\right)}\right) \\ &= \Phi\left(\frac{x_k-1}{2}\right) - \phi\left(\frac{x_{k-1}-1}{2}\right) + 2 \left[\phi\left(\frac{x_{k-1}-1}{2}\right) - \phi\left(\frac{x_k-1}{2}\right)\right] \end{aligned}$$

where  $x_1 = -\infty$  and  $x_K = \infty$ .

### Appendix B.3 Computational details

In both cases 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 (6) 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 C Monte-Carlo Study: Volumes

Table C.1: Monte Carlo Analysis — MMD, Uniform Case — 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 C.2: Monte Carlo Analysis — MMM — Volumes

	Uniform		Normal	
Volume				
N = 100, k = 2	.1348	[.0465, .2435]	.0508	[.0014, .1306]
N = 200, k = 5	.1236	[.0587, .1717]	.0399	[.0185, .0773]
N = 800, k = 20	.1171	[.0980, .1345]	.0220	[.0056, .0326]
N = 20000, k = 30	.1183	[.1151, .1219]	.0152	[.0029, .0231]
True, k = 2	.2228		.0155	
True, k = 200			.0034	
True, k = 500	.1119			
Volume / Volume “Cube”				
N = 100, k = 2	.048	[.028, .066]	.103	[.041, .168]
N = 200, k = 5	.069	[.052, .084]	.164	[.104, .215]
N = 800, k = 20	.087	[.077, .098]	.178	[.093, .228]
N = 20000, k = 30	.085	[.083, .088]	.158	[.064, .208]
True, k = 2	.052		.151	
True, k = 200			.233	
True, k = 500	.087			

## References

- [1] Beresteanu, Arie and Francesca Molinari (2008), “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76(4), 763-814.
- [2] Bontemps, Christian, Thierry Magnac, and Eric Maurin (2011), “Set Identified Linear Models,” *Econometrica*, Forthcoming.
- [3] Frölich, Markus (2006), “Non-parametric Regression for Binary Dependent Variables,” *Econometrics Journal*, 9(3), 511-540.
- [4] Fukuda, Komei, Liebling, Thomas M., and Francois Margot (1997), “Analysis of Backtrack Algorithms for Listing All Vertices and All Faces of a Convex Polyhedron,” *Computational Geometry*, 8, 1-12.
- [5] Hayfield, Tristen and Jeffrey S. Racine (2008), “Nonparametric Econometrics: The np Package.” *Journal of Statistical Software*, 27(5).
- [6] Magnac, Thierry and Eric Maurin (2008), “Partial Identification in Binary Models: Discrete Regressors and Interval Data,” *Review of Economic Studies*, 75(3), 835-864.
- [7] Manski, Charles F. (2003), *Partial Identification of Probability Distributions*, Springer, New York.
- [8] Manski, Charles F. (2007), *Identification for Prediction and Decision*, Harvard University Press, Cambridge, MA.
- [9] Manski, Charles F. and Elie Tamer (2002), “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519-546.
- [10] Menzel, Konrad (2011), “Consistent Estimation with Many Moment Inequalities,” mimeo.
- [11] Tamer, Elie (2010), “Partial Identification in Econometrics,” *Annual Review of Economics*, 2, 167-195.