MISSPECIFICATION IN MOMENT INEQUALITY MODELS: BACK TO MOMENT EQUALITIES?*

MARIA PONOMAREVA AND **ELIE TAMER**

NORTHWESTERN UNIVERSITY

Abstract. Consider the linear *model* $E[y|x] = x'\beta$ where one is interested in learning about β given data on y and x and when y is *interval* measured, i.e., we observe $([y_0, y_1], x)$ such that $P(y \in [y_0, y_1]) = 1$. Moment inequality procedures use the implication $E[y_0|x] \le x'\beta \le E[y_1|x]$. As compared to least squares in the classical regression model, estimates obtained using an objective function based on these moment inequalities do not provide a clear approximation to the underlying unobserved conditional mean function. Most importantly, under misspecification, it is not unusual that no parameter β satisfies the previous inequalities for all values of x, and hence minima of an objective function based on these moment inequalities are typically tight.

We construct set estimates for β that have a clear interpretation when the model is misspecified. These sets are based on moment equality models. We illustrate these sets and compare them to estimates obtained using moment inequality based methods. In addition to the linear model with interval outcomes we also analyze the binary missing data model with a monotone instrument assumption (MIV), we find there that when this assumptions is misspecified, bounds can still be non-empty, and can differ from parameters obtained via maximum likelihood. We also examine a bivariate discrete game with multiple equilibria.

In sum, misspecification in moment inequality models is of a different flavor than in moment equality models, and so care should be taken with the 1) interpretation of the estimates and 2) the size of the 'identified set.'

Date: Original date: March 2008. This version: November 16, 2009.

* We thank the editor and referees for comments that greatly improved the paper. We also thank participants at the Partial Identification conference held in London in March 2008 for comments. Financial Support from the National Science Foundation is gratefully acknowledged.

1. Introduction

It is well known but perhaps often ignored that interpretation of estimates in parametric or semiparametric models is conditioned on well specification of that model. When models are approximations of the true data generating process, estimates provide a particular approximation of the underlying objects of interest. In misspecified moment equality models for example, parameter estimates are ones that minimize some distance of the vector of moments to zero. The interpretation of these parameters from a moment inequality model when the model is misspecified is the subject of this paper.

Recently, there has been a flurry of work in econometrics related to inference in partially identified or incomplete models[1]. These incomplete "models" are based on a strictly smaller set of assumptions than their complete model counterpart in the hopes that conclusions about those parameters obtained using this partial list of assumptions are more robust. The price to pay is that a model with less assumptions is only able to restrict parameters to a non-trivial set - or that the parameter of interest is partially identified. Parametric (or semiparametric) incomplete models consist of a set of maintained assumptions, some are non-testable and others can be rejected in the data. We point out, through a series of simple examples, that care should be taken when interpreting and comparing parameter estimates in partially identified moment *inequality* models. In essence, one needs to first define the object of interest, and then ensure that the objective function used to estimate this parameter allows for a clear interpretation in the presence of misspecification. More crucially, care should be taken when comparing parameters obtained from moment inequality to others obtained using say moment equality models.

In addition, a consequence of misspecification is that partially identified parametric models based on moment inequalities deliver set estimates of the object of interest that are tight (with tight standard errors) when estimated using real data. Under misspecification, it is possible that no parameter vector obeys *all* the moment inequalities and so what is presented, heuristically, is the set of parameter that minimizes a certain criterion function. These small identification "regions" should not

[1] The literature can be classified into identification and estimation in partially identified models. To get a flavor of the identification results see Manski (1995), Manski (2007), Haile and Tamer (2003) and Beresteanu, Molinari, and Molchanov (2008). For a flavor of the estimation and confidence regions literature see Chernozhukov, Hong, and Tamer (2007), Imbens and Manski (2004), Molinari (2003), Andrews and Soares (2009), Romano and Shaikh (2008), Canay (2007) and others.

be interpreted as coming from inference under weaker assumptions[2], rather, this behavior ought to be investigated further. So, even though the motivation behind the partially identified literature has been to gain robustness against certain (or almost all) parametric assumptions that are esoteric to the problem at hand, more robust (but still misspecified) partially identified models are delivering estimates that are as tight as ones obtained with models based on more assumptions. The main reason for obtaining these tight estimates, as we see below in the examples, is that estimates with these moment inequality models are *not* comparable to ones that are obtained with a more complete model since under misspecification these two sets of parameters are estimating different objects (and hence comparing the estimates from these two sets of models is not appropriate).

Finally, we shed light on the choice of objective function in moment inequality models. The choice of the objective function is not motivated by power perspective (or obtaining the smallest set that obeys a confidence property), but by the fact that objective functions, as is well known, provide estimates of different objects when the model is misspecified. Hence, it is important that one thinks of the misspecification issue also when choosing the objective function.

The paper is organized as follows. Section 2 discusses our main ideas in the context of the ubiquitous linear model. We provide a least squares set (LS) that coincides with the identified set for best linear predictor constructed by Stoye (2006) and show that this set contains the set of parameters that provide the best approximation to the underlying conditional mean function. Section 3 discusses a nonparametric missing data model, and then adds a monotone instrumental variable assumption (MIV). There, the set we construct is the one that maximizes a well defined likelihood (that partially identifies the parameter of interest). Interestingly, we show through a simple example that when the MIV is false, the set that one derives based on moment inequalities can be non-empty and differs from the MLE set, while both of these sets coincide if the model is well specified. Section 3 discusses a discrete game example. Throughout, we clarify the issue raised above. Section 4 concludes.

[2]Typically, papers using moment inequalities report tight estimates and confidence regions, and so at first look, these estimates appear as an indicator that the modeling strategy is “succeeding”, in that it is delivering tight estimates with fewer assumptions.

2. Linear Mean Regression with Interval Data

We consider in details the problem of inference on $\beta = (\beta_0, \beta_1) \in \mathcal{R}^2$ in the model

$$E[y|x] = \beta_0 + \beta_1 x \tag{2.1}$$

The analysis can be generalized to mean regressions with k regressors. The model partially identifies β since we assume that we observe $([y_0, y_1], x)$ where $P(y \in [y_0, y_1]) = 1$ and where y_0 is smaller than y_1. We start by considering a complete version of the model in (2.1) above (without censoring) and then introduce censoring and analyze what different partial-identification based methods estimate when this model is not well specified, i.e., when $E[y|x]$ is not linear in x. Throughout the analysis, it is important to keep in mind that the object of interest is the parameter vector β in (2.1) above where if the model is well specified, $\beta_0 + \beta_1 x$ is the true conditional mean function, but if the model is misspecified, $\beta_0 + \beta_1 x$ represents the best linear approximation to the conditional mean function which is the ultimate object of interest[3] in this paper.

In the presence of a random sample of observations on y and x, we can use least squares to obtain consistent estimates of β_0 and β_1. This model is making an important assumption about the relationship between y and x, mainly, that the conditional mean function of y given x is linear in x. Hence, estimating β_0 and β_1 is sufficient to learn this function. It is also well known that if the model above is misspecified, i.e., $E[y|x]$ is not linear in x, then $\beta_0 + \beta_1 x$ is the best linear predictor of y given x under square loss. This line can equivalently be shown to be the straight line that comes closest, in a mean squared sense, to the true conditional mean function, i.e.,

$$\beta = (\beta_0, \beta_1) = \underset{b_0, b_1}{\operatorname{argmin}}\, E\left[\, (E[y|x] - b_0 - b_1 x)^2\, \right]$$

Now, suppose that we do not observe y, rather we observe $[y_0, y_1]$ such that $P(y \in [y_0, y_1]) = 1$. In this situation, there are a few approaches to inference on $\beta = (\beta_0, \beta_1)$. One approach is parametric and is based on making an assumption on the censoring mechanism and the underlying distribution of y (conditional on x). For example, if we have fixed censoring at zero (say), then a normality assumption on the conditional distribution of $y|x$ will allow us to use a likelihood based approach to consistently estimate β (this is the classic Tobit model[4]). Under

[3] Another object of interest can be the best linear predictor under square loss. We do not consider this parameter here. Also, see Goldberger (1991) for reasons why one is interested in the conditional expectation function.

[4] There are other "semiparametric" approaches also, notably the censored LAD model of Powell (1984) and its generalizations to random censoring by Honoré, Khan, and Powell (2002).

misspecification, parameter estimates' interpretation changes and depends on the objective function used. In the Tobit situation, the MLE is the quasi likelihood (White (1994)), and the parameter estimates then are ones that minimize the distance (in the Kullback-Liebler or entropy sense) between the parametric likelihood (model) and the nonparametric likelihood (data). Estimates from these parametric models are more transparent but researchers must pay attention to the sensitivity of their estimates to ad-hoc assumptions[5]. Under general forms of censoring, where all the information that is available is in terms of upper and lower bounds on y (y_0 and y_1), one can still use a parametric assumption on the censoring process (like assume that y is the middle point of $[y_0, y_1]$) and then run least squares. Another approach is semiparametric, where one does not make assumptions on the censoring mechanism, i.e., y can be anywhere in $[y_0, y_1]$ but allow for partial identification (See for example Manski and Tamer (2002) for more on this approach). So, if the model is well specified, this means that for all x,

$$E[y_0|x] \leq \beta_0 + \beta_1 x \leq E[y_1|x] \tag{2.2}$$

The above is a canonical example of a model based on *inequality restrictions*, or *moment inequality model*. One objective function that can be used to make inference on the parameters is the following

$$Q^{mmd}(b_0, b_1) = E_x \left\{ (E[y_0|x] - b_0 - b_1 x)_+^2 + (E[y_1|x] - b_0 - b_1 x)_-^2 \right\} \tag{2.3}$$

where $(u)_+ = (u)1[u \geq 0]$ and similarly for $(u)_-$. Notice that $Q^{mmd}(b_0, b_1) \geq 0$ for all $(b_0, b_1) \in \mathcal{R}^2$. In case the model is well specified, i.e., (2.1) holds, we can easily see that $Q^{mmd}(\beta_0, \beta_1) = 0$. Moreover, with censoring, this objective function is minimized on a nontrivial set of parameters, each of which is a candidate for a conditional mean function. If the model is well specified, the inferential procedure based on minimizing (2.3) will provide the set of (linear) conditional mean functions that are consistent with the data. Hence, one expects that in generic censoring situation (certainly if the censoring is heavy), that the argmin of the above function is not unique. So to do inference in these situations, one can use methods developed

[5]There seems to be a slight confusion in the literature on the difference between robustness and sensitivity. Sensitivity analysis asks the questions of whether *in your data*, the estimates you obtained depend on the underlying assumptions and so sensitivity analysis is linked to a particular data set. For example, if in a data set the least squares estimator is close to the LAD estimator, then we say that the estimates are not sensitive to assumptions on the conditional distribution of $y|x$, but of course, the LAD estimator is more robust (has a bounded influence function in the statistical sense), a statement that one makes *before* confronting the data.

for models that are not point identified. One such approach is Chernozhukov, Hong, and Tamer (2007).

2.1. **Inference under Misspecification.** The problem arises if the model is misspecified, i.e., when the true but unobserved conditional mean of y is not linear in x. Again, the object of interest is still the vector β which is *the best linear approximation* to the (unknown) conditional mean function. Note that minimizers of (2.3) do not have the interpretation of parameters that minimize the squared distance between the true conditional mean function (that is censored) and the set of linear functions. The argmins of (2.3) are parameters that minimize the distance between a certain nonlinear region and the set of linear functions. The nonlinear region is the one bounded above and below by $E[y_0|x]$ and $E[y_1|x]$. This is illustrated in Figure 1 below where the unobserved conditional mean function that is nonlinear, is graphed alongside the upper and lower conditional mean functions. You can see that there does NOT exist a straight line that obeys (2.2) for all x, and hence what comes out of minimizing (2.3) is a parameter vector that minimizes the distance between a line and the upper and lower bounds. See Theorem 2.2 for more on this. Hence, with

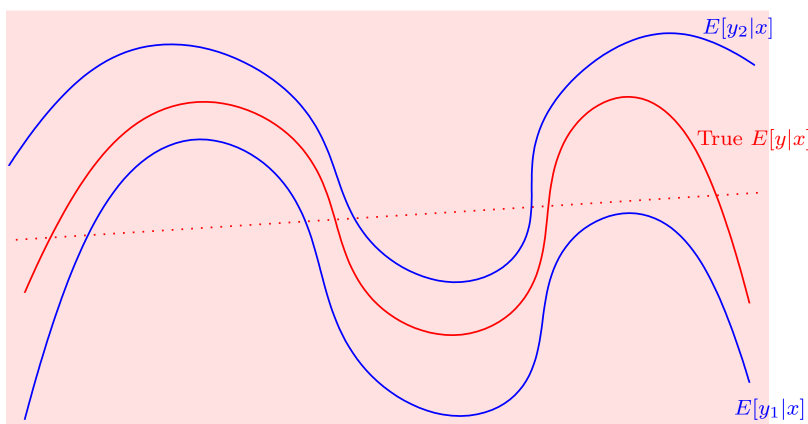


Figure 1. Linear Mean Regression with Interval Data under Misspecification

misspecification, the vector β that minimizes (2.3) does not have the same interpretation as the one that we get from the same misspecified least squares model absence censoring. The moment inequality model in (2.2) applies (2.1) to get (2.2). The argmin of (2.3) is small (and sometimes *unique*, as would be in Figure 1). So, when the model is misspecified, we can get tight estimates even when *no assumptions* are made on the censoring mechanism. However, these estimates do not compare to β, the best linear approximation to the conditional mean function $E[y|x]$. So, we are getting (tight) estimates of a different object.

The next section provides a method to construct estimates of the set of parameters that minimize the squared distance between a linear function and the (unobserved) conditional mean function. We do that by first constructing the set of *all* conditional mean functions that are consistent with the data (this will be made precise below), and then for each of these functions, we find the least squares estimates. This set of estimates is termed the *Least Squares Set* and consists of the set of parameters that are best linear approximations to the (non-identified) conditional mean function.

2.2. **Construction of a Least Squares (LS) Set.** We first describe the method heuristically before providing a practical estimator. Suppose that one can construct all the possible models for $y|x$ that obey the assumptions above, i.e., $P(y \in [y_0, y_1]) = 1$. This a set of conditional probability distributions, $\mathcal{P}_{y|x}$ that is consistent with the data. For each candidate model $P_{y|x} \in \mathcal{P}_{y|x}$, collect the corresponding slope and intercept that minimize the mean squared error by the method of least squares. This is the set of interest. This procedure *completes* the model to obtain a set of fully specified *moment equality models*, each of which leads to an estimate of β that *has the same interpretation* under misspecification that a model without censoring has. This will be illustrated for the linear model in (2.1) above. This intuition of *completing the model* holds in other setups. We will highlight these below in other examples.

In the linear model above, there is a simple procedure that we can use to construct the least squares set estimates for the best linear approximation to the (unobserved) conditional mean function $E[y|x]$. Consider a random variable λ with support on $[0, 1]$ and then construct y_λ as follows:

$$y_\lambda = y_0\lambda + y_1(1 - \lambda)$$

The distribution of λ conditional on all other variables $((y_0, y_1, x))$, F_λ, is left unspecified and it belongs to the set Γ of probability distributions on $[0, 1]$. We then get

$$\beta(F_\lambda) = (\beta_0(F_\lambda), \beta_1(F_\lambda)) = \underset{b_0, b_1}{\text{argmin}}\; E\left[\,(y_\lambda - b_0 - b_1 x)^2\,\right] \tag{2.4}$$

The *least squares* (LS) set is

$$\Theta = \{\beta(F_\lambda) : \quad F_\lambda \in \Gamma\} \tag{2.5}$$

We can derive the asymptotic distribution of the stochastic process $\beta(\cdot)$ indexed by bounded functions as follows. We examine the case of the slope parameter $\beta_1(\cdot)$ for simplicity. Let μ_x and $\widehat{\mu}_x$ be the population mean and sample mean of x respectively. Then,

$$\beta_1(F_\lambda) = \frac{1}{var(x)} E_x \left[(x-\mu_x)\{E(y_1|x) + E(y_0\lambda|x) - E(y_1\lambda|x)\}\right] \tag{2.6}$$

To each distribution function F_λ there corresponds the conditional expectation function $g(u, v_0, v_1) = E(\lambda|x = u, y_0 = v_0, y_1 = v_1)$, which is bounded between 0 and 1, so that $0 \leqslant g(\cdot,\cdot,\cdot) \leqslant 1$. Similarly, any such function g defines some $\beta(F_\lambda)$, so:

$$\beta_1(F_\lambda) \equiv \beta_1(g) = \frac{1}{var(x)} E_x \left[(x-\mu_x)\{E(y_1|x) + E(y_0 g(x,y_0,y_1)|x) - E(y_1 g(x,y_0,y_1)|x)\}\right] \tag{2.7}$$

and its sample analog estimator is

$$\widehat{\beta}_1(g) = \frac{1}{\widehat{var}_n(x)} \widehat{E}_{x,n} \left[(x-\widehat{\mu}_x)\{\widehat{E}_n(y_1|x) + \widehat{E}_n(y_0 g(x,y_0,y_1)|x) - \widehat{E}_n(y_1 g(x,y_0,y_1)|x)\}\right] \tag{2.8}$$

Let $Z_n(g) = \sqrt{n}(\widehat{\beta}_1(g) - \beta_1(g))$ and $\Delta g = g_1 - g_2$. Then, by linearity,

$$|Z_n(g_1) - Z_n(g_2)| = |\tilde{Z}_n(\Delta g)|,$$

where

$$\tilde{Z}_n(\Delta g) = \sqrt{n}\Big(\frac{1}{\widehat{var}_n(x)} \widehat{E}_{x,n} \left[(x-\widehat{\mu}_x)\{\widehat{E}_n(y_0\Delta g|x) - \widehat{E}_n(y_1\Delta g|x)\}\right] - - \frac{1}{var(x)} E\left[(x-\mu_x)\{E(y_0\Delta g|x) - E(y_1\Delta g|x)\}\right]\Big) \tag{2.9}$$

So, we have $\sup\limits_{\|\Delta g\|\leqslant\delta} |\tilde{Z}_n(\Delta g)| = \delta|O_p(1)|$. Therefore, $\forall\varepsilon > 0$ and $\forall\eta > 0$ $\exists\delta > 0$ so that

$$\limsup_{n\to\infty} P\left\{\sup_{\|\Delta g\|\leqslant\delta} |\tilde{Z}_n(\Delta g)| > \eta\right\} < \varepsilon. \tag{2.10}$$

This implies stochastic equicontinuity of $Z_n(g) = \sqrt{n}(\widehat{\beta}_1(g) - \beta_1(g))$ which by standard CLT assumptions, warrants

$$\sqrt{n}(\widehat{\beta}(F_\lambda) - \beta(\lambda)) \overset{\mathrm{L}}{\Longrightarrow} Z(F_\lambda) \tag{2.11}$$

where $Z(\cdot)$ is a gaussian process indexed by $F_\lambda \in \Gamma$.

2.2.1. *Extreme Points.* The *extreme* points[6] of this set can be estimated by exploiting the form of the solution to the optimization problem in (2.4). Starting with the slope, we know that

$$\beta_1(F_\lambda) = \frac{cov(y_\lambda, x)}{var(x)} = \frac{E[(x-\mu_x)y_\lambda]}{var(x)} \tag{2.12}$$

Hence, to obtain the largest and the smallest value of β_1 that belong to Θ, we need to compute the largest and the smallest value of the numerator since the denominator is positive and can be consistently estimated. It is interesting to note that the above bounds are the same as ones derived for best linear predictors under outcome censoring by Stoye (2006). Moreover, the above bounds for cases where x is a vector can be constructed similarly. For example, suppose $x = (x_1, x_2)$. Then, bounds on β_1, the slope coefficient associated with x_1, can be constructed after first premultiplying the regression by the the projection matrix for x_2 hence transforming the regression into one that has a scalar regressor.

Theorem 2.1. *The extreme points of the set* Θ *are the following:*

$$\begin{aligned} \beta_1^{max} &= \frac{1}{var(x)}\left(E\big[(x-\mu_x)y_1 1[(x-\mu_x)\geq 0]\big] + E\big[(x-\mu_x)y_0 1[(x-\mu_x)<0]\big]\right) \\ \beta_1^{min} &= \frac{1}{var(x)}\big(E[(x-\mu_x)y_0 1\big[(x-\mu_x)\geq 0]\big] + E[(x-\mu_x)y_1 1\big[(x-\mu_x)<0]\big] \end{aligned} \tag{2.13}$$

For the intercept, the extreme points are

$$\begin{aligned} \beta_0^{max} &= \frac{1}{var(x)}E\big[(var(x)-(x-\mu_x)\mu_x)1\big[(var(x)-(x-\mu_x)\mu_x)\geq 0\big]\,y_1\big] \\ &+ \frac{1}{var(x)}E\big[(var(x)-(x-\mu_x)\mu_x)1\big[(var(x)-(x-\mu_x)\mu_x)\geq 0\big]\,y_0\big] \\ \beta_0^{min} &= \frac{1}{var(x)}E\big[(var(x)-(x-\mu_x)\mu_x)1\big[(var(x)-(x-\mu_x)\mu_x)\leq 0\big]\,y_0\big] \\ &+ \frac{1}{var(x)}E\big[(var(x)-(x-\mu_x)\mu_x)1\big[(var(x)-(x-\mu_x)\mu_x)< 0\big]\,y_1\big] \end{aligned} \tag{2.14}$$

Proof: Rewrite the slope as follows,

$$\begin{aligned} \beta_1(\lambda) &= \frac{1}{var(x)}E[(x-\mu_x)y_\lambda] \\ &= \frac{1}{var(x)}\big(E[(x-\mu_x)y_\lambda 1\big[(x-\mu_x)\geq 0]\big] + E[(x-\mu_x)y_\lambda 1\big[(x-\mu_x)<0]\big] \end{aligned}$$

[6]We focus on the outer extreme points of the identified for ease of computations. The shape of the identified set in this linear model depends on the support of x. It is a polygon if x has finite support, and it is strictly concave if x is continuously distributed. The set is symmetric around $E(xx')^{-1}E(xy)$ where $y = (y_0 + y_1)/2$. This is all that can be said about the sharp set without further information on the distribution of (x, y_0, y_1).

Hence, the maximum of $\beta_1 \in \Theta$ can be easily obtained by exploiting the fact that $\lambda \in [0,1]$:

(2.15)
$$\beta_1^{max} = \frac{1}{var(x)}\Big(E\big[(x-\mu_x)y_1 1[(x-\mu_x)\geq 0]\big] + E\big[(x-\mu_x)y_0 1[(x-\mu_x)<0]\big]\Big)$$
$$\beta_1^{min} = \frac{1}{var(x)}\Big(E[(x-\mu_x)y_0 1\big[(x-\mu_x)\geq 0]\big] + E[(x-\mu_x)y_1 1\big[(x-\mu_x)<0]\big]$$

This can be easily estimated consistently with the observed data. Similarly to the slope, the intercept extreme points can be estimated as follows.

(2.16)
$$\begin{aligned}\beta_0(\lambda) &= \mu_{y_\lambda} - \beta_1(\lambda)\mu_x \\ &= E\big[\big(1-\frac{(x-\mu_x)}{var(x)}\mu_x\big)y_\lambda\big] = E\big[\big(\frac{var(x)-(x-\mu_x)\mu_x}{var(x)}\big)y_\lambda\big]\end{aligned}$$

It easy to see that the desired result for the intercept follows. □

As we can see, the above theorem provides the extreme points of the set Θ in (2.5). Each parameter in this set represents a best linear approximation to *some* conditional expectation function that lies between the upper and lower conditional expectation functions. So, any two conditional expectations $E[y_{\lambda_1}|x]$ and $E[y_{\lambda_2}|x]$ that are consistent with the model are treated equally in that $\beta(\lambda_1)$ and $\beta(\lambda_2)$ belong to Θ.

2.3. **The MMD Parameter.** The previous section derives bounds on the best approximation to the conditional expectation function that is unobserved but is known to lie between observed upper and lower bounds. Here, we provide a link between between those best approximations to the conditional expectation and the minimizers of the MMD objective function

Theorem 2.2. *Let*

$$(\beta_0^{mmd}, \beta_1^{mmd}) \in \arg\min_{b_0,b_1} Q^{mmd}(b_0,b_1)$$

Define the function Q as follows:

(2.17)
$$Q(b_0,b_1,\lambda_0,\lambda_1) = E\big[(E[y_0|x]+\lambda_0(x)-b_0-b_1x)^2 + (E[y_1|x]-\lambda_1(x)-b_0-b_1x)^2\big]$$

Then,

$$(\beta_0^{mmd}, \beta_1^{mmd}, \lambda_0^*(x), \lambda_1^*(x)) \in \arg\min\ Q(b_0,b_1,\lambda_0,\lambda_1)$$

where $\lambda_0^*(x) = (\beta_0^{mmd} + \beta_1^{mmd} x - E[y_0|x])_+$ *and* $\lambda_1^*(x) = (E[y_1|x] - \beta_0^{mmd} - \beta_1^{mmd} x)_+$.

Proof: A direct way to prove this is to show that the MMD objective function can be written as in (2.17). Fix b_0, b_1. Then the optimal choice of λ_0 and λ_1 is

$$\lambda_0(b_0, b_1) = (b_0 + b_1 x - E[y_o|x])_+ \text{ and } \lambda_1(b_0, b_1) = (E[y_1|x] - b_0 - b_1 x)_+$$

which is exactly $Q^{mmd}(b_0, b_1)$. Therefore any solution to $\min Q^{mmd}(b_0, b_1)$ is also a solution to $\min Q(b_0, b_1, \lambda_0, \lambda_1)$ with λ_0 and λ_1 as above. The reverse of this claim is also true. □

The minimizers of the MMD objective function also minimize the sum of two approximation errors in (2.17). The first one measures the squared distance between the upper conditional expectation $E[y_1|x]$ and a partly linear function and the second piece is similar. So, as we can see, the minimand of the MMD objective function minimizes the sum of squared approximations in (2.17) and not a least squares like objective function and naturally then the interpretations of both under misspecification differ.

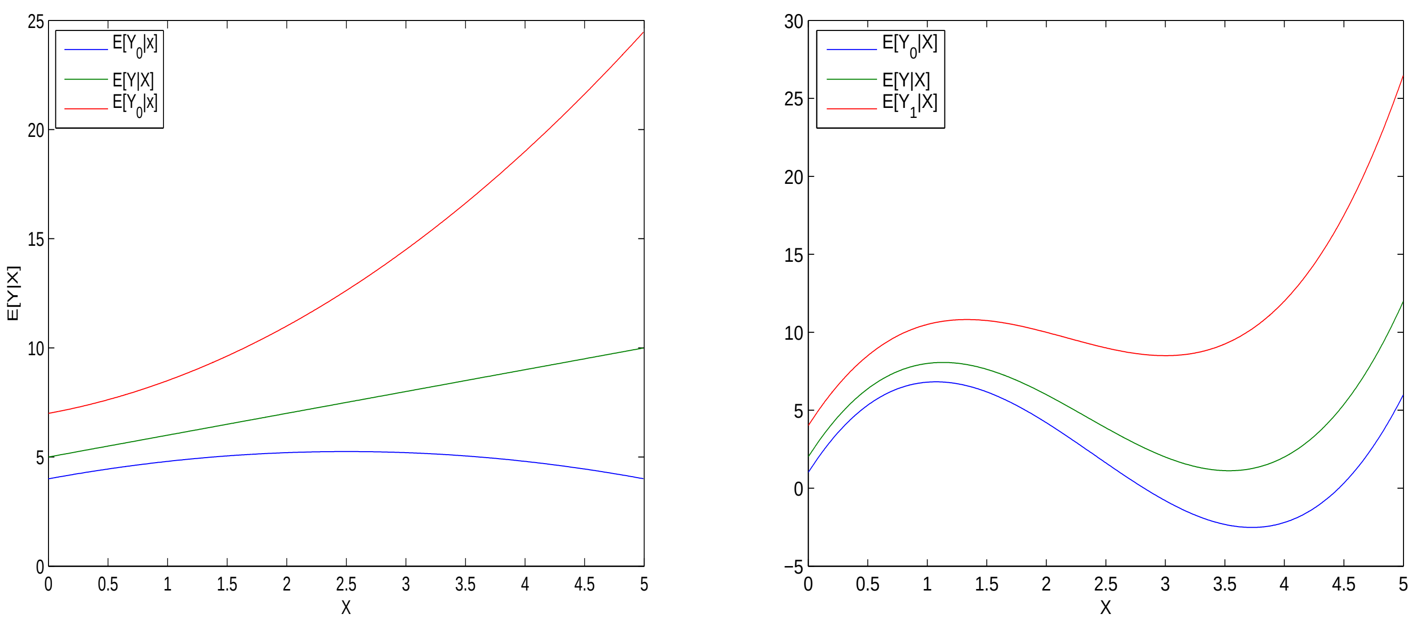
What not to compare: In principle, the LS set and the MMD sets are both interesting objects to report. For example, the LS set is attractive since it contains parameters that are directly comparable to the parameter one would get if no censoring occurs. This is relevant if the objective of the empirical exercise is the underlying best approximation to the conditional mean function. This is exactly what one would be after when running a *linear regression* absence censoring. The MMD set on the other hand contains the truth when the model is well specified, but otherwise, this set contains parameters that are defined as in Theorem 2.2. Also, when the MMD objective yields the empty set, then, we know that the model is misspecified and that the conditional expectation is not linear. On the other hand, a non-empty MMD set is not indicative that the model is well specified.
With misspecification, *the two sets can have an empty intersection* and are always different. Estimates obtained using a complete model (through some least squares procedure such as regressing the midpoint of y_0 and y_1 on x) should not be compared to estimates from the MMD set since each is estimating a different object. Care should be taken with the interpretation, and especially testing, in moment inequality models.

2.4. **Monte Carlo Experiments.** For purposes of exposition, we ran a set of Monte Carlo experiments that illustrates the identified set that one obtains using MMD, which relies on well specification, to the least squares method we propose for the linear model. The calculations are done on the population level. In particular, for the MMD model, we plot the contour set that minimizes (2.3) above. The minimum of the objective function is strictly positive if the model is misspecified. First, for a well specified model, the conditional mean of $y|x$ is a linear function of x where x is a uniform random variable on $[0, 5]$. We generate the two censoring variables y_0 and y_1 by adding a positive function of x to y and subtracting another positive function of x to y respectively.

$$
\begin{aligned}
&E[y|x] = 5 + x \quad \textbf{well specified model} \\
&E[y|x] = 5 + (x-2)^3 - x^2 \quad \textbf{misspecified model}
\end{aligned}
\tag{2.18}
$$

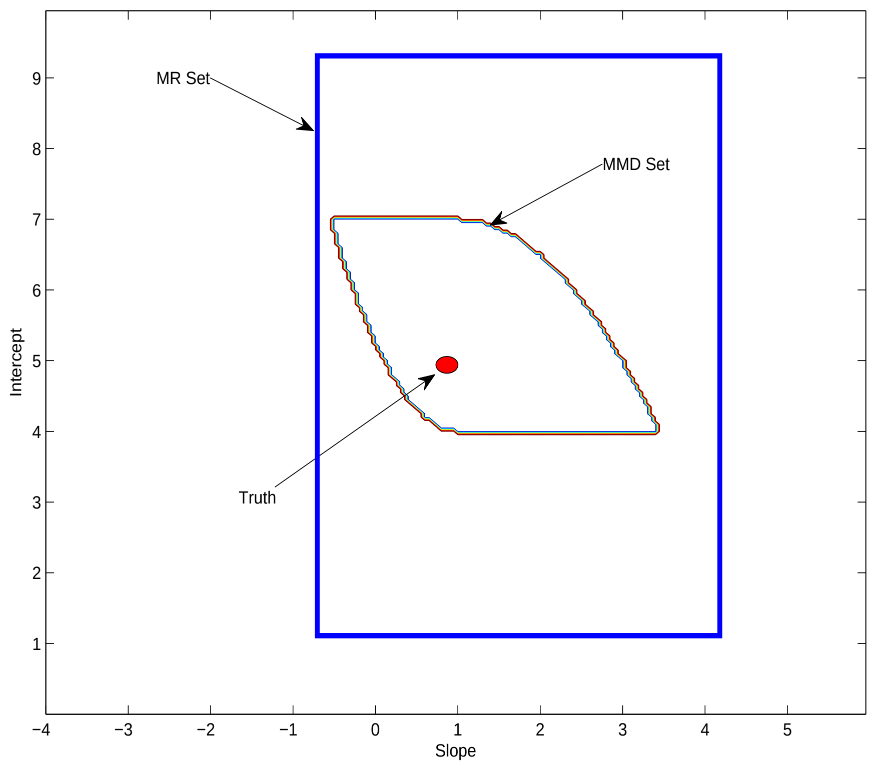
We make the censoring depend on x. The left hand panel in Figure 2 below provides a graph of the conditional mean functions in the case of a well specified model along with the conditional expectation of the upper and lower envelopes $E[y_1|x]$ and $E[y_0|x]$. The right hand panel of the same figure provides an example of a model in which the conditional expectation function is misspecified along with upper and lower envelopes on it.

FIGURE 2. Well Specified (LHS) and Misspecified Model (RHS)



There will be a set of lines that will fit between the upper and lower curves in the LHS display. The set of slopes and intercepts corresponding to this set of lines is the identified set. However, from the second display, we see that there will be NO lines that fit entirely (at least in $[0, 5]$) between the upper and lower envelopes. So, here, one would expect that an MMD estimator would result in a unique parameter. In Figure 3, we first plot the MMD set and the LS set in the well specified case.

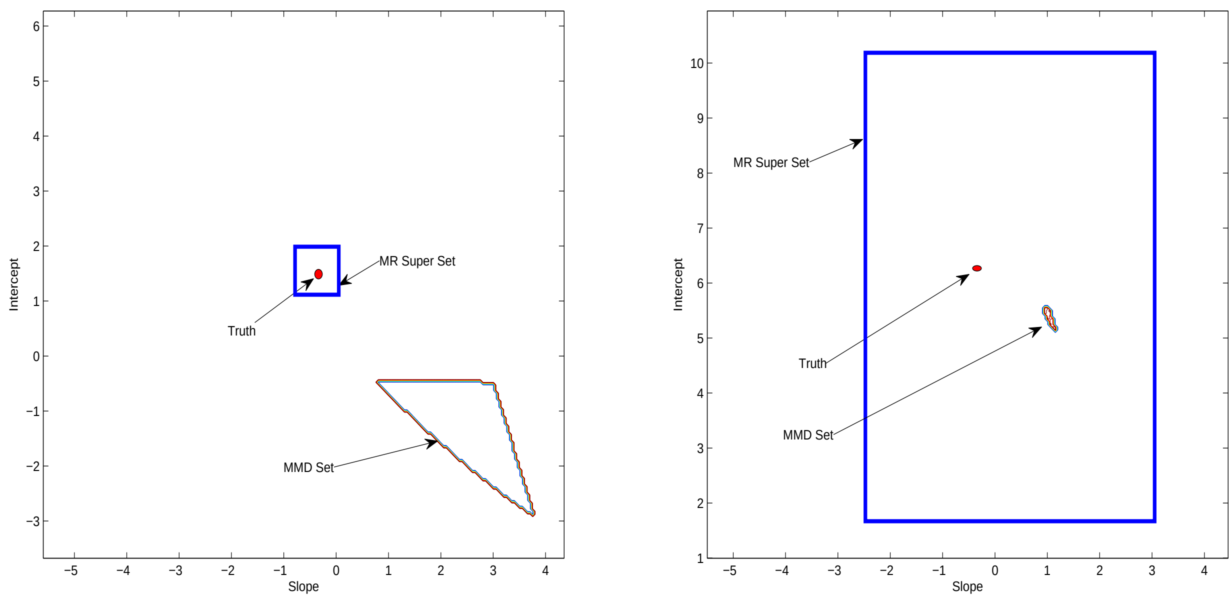
FIGURE 3. LS vs MMD Sets in the Well Specified Case



As one can see, the true parameter is contained in the MMD set which is smaller and contained within the LS set. The MMD is the set at which the population MMD is equal to zero. It contains the set of lines that obey the model, i.e., that fit within the upper and lower bound functions. On the other hand, the extreme points of the LS set are computed as we did above, and this set contains the best linear approximations to *all* conditional expectations, *both linear and nonlinear* that fit between $E[y_0|x]$ and $E[y_1|x]$. So, obviously, this set is bigger than the MMD set in the well specified case. This set will further shrink to the MMD set if we keep parameters (b_0^*, b_1^*) such that $E\Big[\,(E[y_\lambda|x] - b_0^* - b_1^* x)^2\,\Big] = 0$

In the case of misspecification, I plot, in Figure 4, the LS set, and the MMD set along with the true parameter vector β that is the best linear approximation to the (unknown) conditional expectation function. In the LHS display, we see the LS and

FIGURE 4. LS and MMD in the Misspecified Case



the MMD set that correspond to the misspecified model in Figure 3 above. We see that the LS set contains the best linear approximation to the true (but unobserved) conditional expectation. The MMD set corresponds to the set of parameters where the MMD function is minimized, which in this case is strictly positive. This set does not contain the best linear approximation to the conditional expectation. On the other hand, the LS (super) set is tight and does (and will always) contain the best linear approximation to the conditional expectation. So, as is argued above, the LS procedure and the MMD procedure are estimating different objects and so comparing the MMD set to *any* vector in the LS set can be problematic.

Notice now the display in the RHS of Figure 4. This corresponds to another design and is produced here to shed light on an interesting case. The important feature that this plot highlights is how small the MMD set is. In fact, with minor changes in the design, it can be reduced to basically a point. If an empirical worker uses the MMD objective function, s/he will obtain tight estimates *even in the presence of censoring* which runs contrary to what one would expect in models with interval-censored outcomes. The MMD set estimates its parameter well which, because of misspecification, is not comparable to β. Notice that the LS set which estimates the set of best linear approximations to a conditional expectation function that obeys the model and is consistent with the data is wide.

3. More Examples

We provide other illustrative examples of the issues raised above. The first example examines a nonparametric partially identified model where no testable assumptions are made. There, we can see that the MMD procedure as well as the LS procedure give the same answers. This is because in nonparametric models, there is no scope for misspecification. Then, we add a monotone instrumental variable assumption and we characterize cases when simple models with inequality restrictions can give different estimates as compared to maximum likelihood based estimates. The second example we give is one of a simple game with multiple equilibria. Again, here we provide inference procedures that are based on the likelihood (as opposed on some moment inequalities) that have a natural interpretation under misspecification.

3.1. **Nonparametric Missing Data Model.** Here, consider the problem of learning the mean β of a binary variable y, but we only observe y when $z = 1$. Hence, we have

$$\begin{aligned} \beta = P(y=1) &= P(y=1|z=1)P(z=1) + P(y=1|z=0)P(z=0) \\ &= P(y=1|z=1)P_1 + \alpha(1-P_1) \end{aligned} \tag{3.1}$$

In the complete model where y is also observed when $z = 0$, one can use maximum likelihood to get $\beta = P(y = 1)$ and hence that the MLE for β is the sample choice probability. In addition, this is a well specified model since it is nonparametric. Now, we only observe y when $z = 1$, then, one possible parametrization of the likelihood is in terms of two parameters: β, the parameter of interest, and $\alpha = P(y = 1|z = 0)$ which is another parameter. The observed data is $\big(y_{z=1}, z\big)$ and hence the MLE maximizes the following likelihood

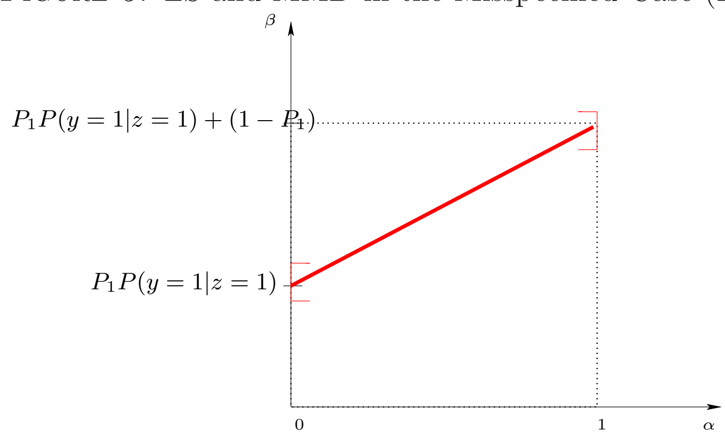
$$L(\beta,\alpha) = E\big[y \log \frac{\beta - (1-P_1)\alpha}{P_1} + (1-y)\log \frac{P_1 - \beta + (1-P_1)\alpha}{P_1}\big]$$

Maximizing this likelihood is meaningful under misspecification also since it provides the parameters that minimize the entropy distance (or the Kullback-Leibler Information Criterion) between the true observed data density and the precited model density. Again, here, the model is nonparametric and so well specified, but the argmax of the likelihood is not unique, or the parameter vector (α, β) is not point identified. Note here that we do not use moment inequalities. Rather, we use a parametrization for the likelihood of the *observed* data (there are other parametrizations) and so the maximizers of the above function L are the set of (b, a) such that

$$\Theta = \{(b, a) \in [0, 1]^2 : \quad P(y = 1|z = 1) = \frac{b - a(1 - P_1)}{P_1}\}$$

This set is the interval graphed in Figure 5.

Figure 5. LS and MMD in the Misspecified Case (2)



In particular, it maps into the following interval for β

$$\beta \in \left[P_1 P(y = 1|z = 1),\ P_1 P(y = 1|z = 1) + 1 - P_1\right]$$

which is exactly the Manski worst-case bound for β. This case is not very interesting again since the model is nonparametric and so there is no scope for misspecification.

3.2. **Adding a MIV assumption:** Suppose we make an *Monotone Instrumental Variable* (MIV) assumption similar to one made in Manski (2003). Mainly, we assume that there exists a binary variable v such that

$$P(y = 1|v = 1) \geq P(y = 1|v = 0)$$

This is in contrast to the (usual) IV assumption which would require statistical independence between y and v. Consider the following definitions.

(3.2)

$$\begin{aligned} P_1 &= P(z=1) \quad P_0 = P(z=0) \\ \pi_0 &= P(v=0|z=0) \quad \pi_1 = P(v=1|z=0) \\ q_0 &= P(z=0|v=0) \quad q_1 = P(z=0|v=1) \end{aligned}$$

$$\Gamma_0 = P(y=1|v=0,z=1)P(z=1|v=0) \quad \Gamma_1 = P(y=1|v=1,z=1)P(z=1|v=1)$$

$$\beta = P(y=1), \alpha = P(y=1|v=0,z=0), \gamma = P(y=1|v=1,z=0)$$

Tight bounds on $\beta = Pr(y=1)$ under the MIV assumption have been derived by Manski and Pepper (2000) and Manski (2007):

(3.3)

$$\Theta_{MIV} = [P(v=0)\Gamma_0 + P(v=1)\max\{\Gamma_0,\Gamma_1\}, P(v=0)\min\{\Gamma_0+q_0,\Gamma_1+q_1\} + P(v=1)(\Gamma_1+q_1)]$$

If the set Θ_{MIV} is empty, then the MIV assumption does not hold. However, the reverse in general is not true: nonempty Θ_{MIV} does not necessarily imply that the monotonicity assumption holds. Therefore, in the case of the misspecified model, i.e. the model where the DGP violates the monotonicity assumption, the above inference procedure may produce nonempty (and sometimes quite tight) bounds that do not cover the true value of β. It is unclear what is the interpretation of a nonempty Θ_{MIV} in this case.

An alternative approach would be to construct the maximum likelihood set (MLS) for the parameter β that allows for a meaningful interpretation even when the MIV assumption does not hold. To do so, we consider the following constrained likelihood maximization problem:

$$\max_{\beta,\mu,\nu} E\left[y\log\left(\frac{\beta - P_0(\pi_0\mu+\pi_1\nu)}{P_1}\right) + (1-y)\log\left(1 - \frac{\beta - P_0(\pi_0\mu+\pi_1\nu)}{P_1}\right)\right]$$

$$\text{s.t.} \quad \begin{cases} \Gamma_1 + q_1\mu \geq \Gamma_0 + q_0\nu \\ 0 \leq \beta,\mu,\nu \leq 1 \end{cases}$$

We see that this gives us the following bounds for β:

(3.4)

$$\Theta_{MLS} = [P_1\min\{P(y=1|z=1), \frac{1-P_0M_0}{P_1}\} + P_0M_0, P_1\min\{P(y=1|z=1), \frac{1-P_0M_0}{P_1}\} + P_0M_1]$$

where

$$M_0 = \max\{0, \pi_0\frac{\Gamma_0-\Gamma_1}{q_1}\} \text{ and } M_1 = \min\{1, \pi_0 + \pi_1\frac{\Gamma_1-\Gamma_0+q_1}{q_0}\} \tag{3.5}$$

If the underlying model is well specified, then $\Theta_{MLS} = \Theta_{MIV}$, that is, the two sets coincide. Alternatively, if the model is misspecified, i.e. the Monotone Instrumental Variable assumption does not hold, the MLS set contains the parameter values that minimize the Kullback-Liebler divergence between the data and a model with imposed MIV restriction. So, when using the monotone instrumental variable assumption, three scenarios can happen. (a) the MIV assumption is well specified, and so here, $\Theta_{MLS} = \Theta_{MIV}$. (b) The MIV is misspecified, and both Θ_{MIV} and Θ_{MLS} are empty, and most interestingly (c) where MIV is misspecified but typically $\Theta_{MIV} \neq \Theta_{MLS}$ and both do not violate the no-information bounds. Typically, both of these sets can be really tight, and hence under misspecification, both (b) and (c) can lead to the 'crossing' problem when constructing Θ_{MIV} (see Manski and Pepper (2000)). It is also common that under (c) both sets are disjoint and hence care should be taken as to which estimates to report. It is not clear what the meaning of the Θ_{MIV} is when the MIV assumption is misspecified. The next simple examples provides cases that illustrate (c).

Examples: Consider the following model:

$$
\begin{aligned}
&P(y = 1|v = 0, z = 0) = P(y = 1|v = 0, z = 1) = 3/4\\
&P(y = 1|v = 1, z = 0) = P(y = 1|v = 1, z = 0) = 1/6\\
&P(z = 1|v = 0) = 1/4, P(z = 1|v = 1) = 5/6\\
&P(v = 1) = 1/5
\end{aligned}
$$

Here $\beta = 0.63$, and the (nonparametric) no-information bound is $[.17, .81]$ and of course β belongs to the no-information bound (or nonparametric bound). This model violates the monotonicity assumption. The set Θ_{MIV} is not empty. In particular, $\Theta_{MIV} = [0.19, 0.31]$. We notice that these bounds are somewhat tight, and that they do not cover the true β, so that $\beta \notin \Theta_{MIV}$. The MLS set is $\Theta_{MLS} = [0.35, 0.78]$. So, this is a case where empirical researchers obtain different results (Θ_{MIV} and Θ_{MLS} are disjoint) depending on the estimation method and there is no reason to reject the MIV assumption since both of these sets lie in the no-information bound. However, the Θ_{MLS} contains the set of β's that minimize the distance between the true observed data likelihood and the model likelihood (that is the nonparametric likelihood under the monotonicity assumption).

A more striking case emerges in the example above when $P(z = 1|v = 0) = 2/5$, keeping other probabilities without changes, so that the parameter value remain the same: $\beta = 0.63$ while the no information bound is now $[.27, .78]$. Again, Θ_{MIV} is nonempty and gives us almost a point-like identification region $\Theta_{MIV} = [0.30, 0.31]$. The MLS set, Θ_{MLS} is also quite tight, $\Theta_{MLS} = [0.73, 0.75]$. There is here a striking difference between the information the two sets provide. The message of these stylized example is that under misspecification, care should be taken in terms of interpretation of the results and the approach one takes to estimate parameters of interest.

3.3. **A Simple Bivariate Game.** Consider the following bivariate game:

$$\begin{aligned} y_1 &= 1[\alpha_1 + \Delta_1 y_2 - \epsilon_1 \geq 0] \\ y_2 &= 1[\alpha_2 + \Delta_2 y_1 - \epsilon_2 \geq 0] \end{aligned} \tag{3.6}$$

As usual, we assume that we observe both y_1 and y_2 and that both Δ_1 and Δ_2 are strictly negative and $(\epsilon_1, \epsilon_2) \sim F_\theta$. Let the parameter of interest be $\gamma = (\theta, \Delta_1, \Delta_2, \alpha_1, \alpha_2)$. It is well known that the game above admits multiple equilibria when

$$(\epsilon_1, \epsilon_2) \in [\alpha_1 + \Delta_1, \alpha_1] \times [\alpha_2 + \Delta_2, \alpha_2] \equiv \Psi$$

In particular, there, we have three equilibria 2, $(1, 0)$ and $(0, 1)$, in pure strategies, and

$$(p_1, p_2) = \left(-\tfrac{\epsilon_1+\alpha_1}{\Delta_1}, -\tfrac{\epsilon_2+\alpha_2}{\Delta_2}\right)$$

in mixed strategies where p_1 is player 1's probability of playing 1 and similarly for player 2. So, the choice probabilities are given by

(3.7)

$$\Pr(1,1) = F_\theta(\alpha_1 + \Delta_1, \alpha_2 + \Delta_2) + \int_\Psi \big(-\frac{\epsilon_1 + \alpha_1}{\Delta_1}\big)\big(-\frac{\epsilon_2 + \alpha_2}{\Delta_2}\big) S_3(\epsilon_1, \epsilon_2) dF_\theta(\epsilon_1 \epsilon_2)$$

$$\Pr(1,0) = \int_{\Psi_{(1,0)}} dF_\theta(\epsilon_1 \epsilon_2) + \int_\Psi S_1(\epsilon_1, \epsilon_3) dF_\theta(\epsilon_1 \epsilon_2) + \int_\Psi \big(-\frac{\epsilon_1 + \alpha_1}{\Delta_1}\big)\big(1 + \frac{\epsilon_2 + \alpha_2}{\Delta_2}\big) S_3(\epsilon_1, \epsilon_2) dF_\theta(\epsilon_1 \epsilon_2)$$

$$\Pr(0,1) = \int_{\Psi_{(0,1)}} dF_\theta(\epsilon_1 \epsilon_2) + \int_\Psi S_2(\epsilon_1, \epsilon_3) dF_\theta(\epsilon_1 \epsilon_2) + \int_\Psi \big(1 + \frac{\epsilon_1 + \alpha_1}{\Delta_1}\big)\big(-\frac{\epsilon_2 + \alpha_2}{\Delta_2}\big) S_3(\epsilon_1, \epsilon_2) dF_\theta(\epsilon_1 \epsilon_2)$$

where

$$S_i(\epsilon_1, \epsilon_2) \in [0, 1] \quad \text{for } i = 1, 2, 3 \text{ and } S_1(\epsilon_1, \epsilon_2) + S_2(\epsilon_1, \epsilon_2) + S_3(\epsilon_1, \epsilon_2) = 1$$

$$\Psi_{(0,1)} = \{(\epsilon_1, \epsilon_2) \in \mathcal{R}^2 : (\epsilon_1 \leq \alpha_1, \epsilon_2 \geq \alpha_2) \cap (\epsilon_1 \leq \alpha_1 + \Delta_1; \alpha_2 + \Delta_2 \leq \epsilon_2 \leq \alpha_2)\}$$

and similarly for $\Psi_{(1,0)}$. The functions S_i above are selection mechanisms that depend on unobservables and hence are general. They do not have a structural interpretation, but are functions that *complete* the model. The above provide a set of choice probabilities predicted by the model and the above then constitute a set of moment *equalities* that can be used to do inference on γ in the presence of the functions S_i.

One approach to inference in the model above is to exploit the fact that the S functions are probabilities and hence by monotonicity of the choice probabilities in S obtain predicted upper and lower bounds on observed choice probabilities. This will transform the model into one with *inequality restrictions* on regressions where an MMD like approach can be used to estimate its parameters. For example, an implication of the above model is the following set of moment inequality restrictions (abstracting from mixed strategies for simplicity)

$$
\begin{aligned}
\Pr(1,1) &= F_\theta(\alpha_1+\Delta_1, \alpha_2+\Delta_2) \\
\int_{\Psi_{(1,0)}} dF_\theta(\epsilon_1\epsilon_2) \le \Pr(1,0) &\le \int_{\Psi_{(1,0)}} dF_\theta(\epsilon_1\epsilon_2) + \int_{\Psi} dF_\theta(\epsilon_1\epsilon_2) \\
\int_{\Psi_{(0,1)}} dF_\theta(\epsilon_1\epsilon_2) \le \Pr(0,1) &\le \int_{\Psi_{(0,1)}} dF_\theta(\epsilon_1\epsilon_2) + \int_{\Psi} dF_\theta(\epsilon_1\epsilon_2)
\end{aligned}
\tag{3.8}
$$

Again, the set of parameter that satisfy the above inequality restrictions when the model is well specified is a superset that contains the identified set. However, in the presence of misspecification, this approach estimates a set of parameters that is not necessarily the set of interest. In fact, under misspecification, the inequalities above will not *all* be satisfied by *any* parameter vector[7].

Partially Identified Maximum Likelihood: The ML Set

Define the maximum likelihood set (or ML set) as follows. Let the data given choice probability vector be equal to $\mathcal{P} = (\Pr(1,1), \Pr(1,0), \Pr(0,1), \Pr(0,0))$. In addition, let the predicted vector of choice probabilities as a function of θ and the $S = (S_1, S_2, S_3)$'s given in (3.7)) be $\mathcal{P}_{\gamma,S}$. Hence, define the set of parameters consistent with the observed probabilities as

$$\chi = \{(\gamma, S) : \; P\big(\mathcal{P} = \mathcal{P}_{\gamma,s}\big) = 1\}$$

This set can also be defined as the argmax of the corresponding likelihood

$$\chi = \; \text{argmax}_{\gamma',S}\, E[\log(\mathrm{P}_{\gamma',S})]$$

Now, the attractiveness of the above approach as compared to the one that uses inequalities (and hence MMD) is that under misspecification, the maximum likelihood

[7]It is irrelevant whether one uses the above non-sharp inequalities to conduct inference or the sharp inequalities of Beresteanu, Molinari, and Molchanov (2008).

set

$$\chi = \text{ argmax}_{\theta,S}\, E[\log(\text{P}_{\theta,S})]$$

minimizes the distance between the true choice probabilities ($\mathcal{P}$) and the predicted ones. This distance again is the well known entropy or KLIC. Hence, a more interpretable approach is to use the maximum likelihood objective function as a function of both γ and the vector of functions $S(\cdot)$. In case the model is well specified, this approach delivers the identified set (and hence is sharp), and on the other hand this has a clear meaning when the model is misspecified. The negative of this approach is that one needs to deal with an infinite dimensional parameter. For a similar approach in a difference context, see Honoré and Tamer (2006).

This approach of completing the model arises since the reason why one obtains inequality restrictions in these games is the fact that one is not willing to model the selection mechanism. Completing the model to obtain a standard likelihood (a semiparametric one here since it depends on unknown functions), or to obtain a moment *equality* model will lead to inference that provides an interpretation when the underlying model is misspecified. Of course, both ensuing MLE or GMM model will likely be partially identified and hence one needs to use inference procedure in moment equality models that are robust to non-point identification (of both γ and the vector of functions S in this case).

4. Conclusion

This paper makes a simple point. Under misspecification, the parameters that one obtains using some moment inequalities might not be ones that researchers are interested in. For example, in linear models with data on both outcomes and regressors, least squares estimates have a clear meaning under misspecification. These are called the least squares quasi-true parameters, or the parameters that provide the best approximation to the conditional mean function. With interval measurement on outcomes, the model implies a set of moment inequalities, that under misspecification, will not provide information about these quasi-true parameters. This "problem" of interpretation does not arise in moment equality models since the meaning of these estimates are basically defined through say the objective function which basically minimizes some particular distance of the vector of moment conditions from zero. With moment inequalities, the interpretation is not as clear and hence care should be taken in interpreting set estimates that are obtained from these models. More importantly, in moment inequalities where one expects the model parameters to be partially identified with often times wide bounds (depending on the assumptions), under misspecification, moment inequality based models tend to provide *tight* estimates of the identified set. This can be a *direct result* of the misspecification and therefore cannot be viewed as an indicator that the underlying model contains a lot of information about the true but partially identified parameter.

References

Andrews, D. W. K., and G. Soares (2009): "Inference for Parameters Defined by Moment Inequalities using Generalize Moment Selection," forthcoming, Econometrica.

Beresteanu, A., F. Molinari, and I. Molchanov (2008): "Sharp Identification Regions in Games," Working Paper.

Canay, I. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Northwestern Working Paper.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Goldberger, A. (1991): *A Course in Econometrics*. Harvard University Press.

Haile, P., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Honoré, B., S. Khan, and J. Powell (2002): "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 64, 241–278.

Honoré, B., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Manski, C. F. (1995): *Identification Problems in the Social Sciences*. Harvard University Press.

——— (2003): *Partial Identification of Probability Distributions*, Springer Series in Statistics. Springer, New York.

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, forthcoming.

Manski, C. F., and J. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–547.

Molinari, F. (2003): "Partial Identification of Probability Distributions with Misclassified Data," Cornell University Working Paper.

Powell, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, pp. 303–325.

Romano, J., and A. Shaikh (2008): "Inference for the Identified Set in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

Stoye, J. (2006): "Bounds on Generalized Linear Predictors with Incomplete Outcome Data," *Reliable Computing*, 13, 293–302.

White, H. (1994): *Estimation, Inference and Specification Analysis*, vol. 22 of *Econometric Society Monographs*. Cambridge University Press.