University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 0920

Bounds on Counterfactual Distributions Under Semi-Monotonicity Constraints

Stefan Boes

December 2009

Bounds on Counterfactual Distributions Under Semi-Monotonicity Constraints

Author's address: Stefan Boes
                       E-mail: boes@sts.uzh.ch

Bounds on Counterfactual Distributions Under
Semi-Monotonicity Constraints[*]

Stefan Boes[†]

December 29, 2009

Abstract

This paper explores semi-monotonicity constraints in the distribution of potential outcomes, first, conditional on an instrument, and second, in terms of the response function. The imposed assumptions are strictly weaker than traditional instrumental variables assumptions and can be gainfully employed to bound the counterfactual distributions, even though point identification is only achieved in special cases. The bounds have a simple analytical form and thus have much practical relevance in all instances when strong exogeneity assumptions cannot be credibly invoked. The bounding strategy is illustrated in a simulated data example and applied to the effect of education on smoking.

JEL Classification: C14, C21, C25

Keywords: Nonparametric bounds, treatment effects, causality, endogeneity, instrumental variables, policy evaluation.

[†]Address for correspondence: University of Zurich, Socioeconomic Institute, Zuerichbergstrasse 14, CH-8032 Zurich, Switzerland, phone: +41 44 634 2301, email: boes@sts.uzh.ch.; Harvard University, Institute for Quantitative Social Science, 1730 Cambridge Street, Cambridge, MA 02138, USA, phone: +1 617 649 9632, email: sboes@iq.harvard.edu.

1 Introduction

A major concern about traditional instrumental variables (IV) methods is the validity of the imposed exogeneity assumption. For example, in the standard linear model, a valid instrument is assumed to be correlated with the endogenous regressor and uncorrelated with the outcome error. In the treatment literature, the distribution of potential outcomes is assumed to be independent of the instrument, perhaps conditional on covariates. Full independence is sometimes replaced by the weaker equality of mean outcomes, but in any case, the credibility of the imposed exogeneity assumption is inevitably subject to dispute since it cannot be formally tested using the empirical evidence alone.

The aim of this paper is to explore the identifying power of modest functional form assumptions. The restrictions are imposed on the potential outcome distribution, first, conditional on an order restriction in one covariate (which will be denoted the instrument), and second, in terms of the response function. For example, the equality of these distributions over different values of the instrument, as imposed by common IV models, is replaced by a sequence of weak inequalities, allowing for global extremum points. In accordance with the previous literature (e.g., Manski and Pepper 2000, 2009), the conditions will be referred to as monotonicity and/or semi-monotonicity assumptions, depending on the shape of the distribution. The identification strategy partially addresses Manski and Pepper's (2000) call for extension of their monotone instrumental variables approach to stochastic dominance.

Three main implications can be drawn from the analysis. First, although point identification of the counterfactual distributions is generally not feasible, the derived bounds are sharp given the assumptions. Thus, without imposing additional stucture on the data-generating process, no improvements over the stated bounds can be obtained. Second, if the instrument is the treatment indicator itself, then the bounds have a very simple analytical form, it is straightforward to calculate their sample values, and to conduct inference using known methods. Third, the paper avoids any distributional, independence, or exclusion restrictions. Hence, the derived bounds have important practical relevance in all cases when such assump-

tions cannot be credibly invoked.

While the assumptions are all motivated in the context of ordinal outcomes as observed, for example, in the evaluation of future economic developments (e.g., Broz and Plouffe 2010), records about educational attainment (e.g., Ermisch and Francesconi 2001), or measures of individual health (e.g., Wildman and Jones 2008), the proposed methods are equally applicable to continuous and mixed discrete/continuous outcomes. The distinction is inconsequential for the partial identification results and the derivation of bounds, but it has implications for the estimation and inference methods. The former is subject of this paper, issues regarding the latter will only be briefly discussed at the end.

The partial identification of counterfactual quantities under weak monotonicity constraints has been initiated in the literature by Manski (1995, 1997) and Manski and Pepper (2000). Subsequent research further developed and applied the idea of monotone instrumental variables (MIV), and monotone treatment response (MTR), in the context of wage regressions (Blundell *et al.* 2007, Gonzáles 2005), health care utilization (Gerfin and Schellhorn 2006, Kreider and Hill 2009), reporting errors (Kreider and Pepper 2007, 2008), and the intergenerational transmission of schooling (de Haan 2008). Okumura and Usui (2007) combine the concave-MTR assumption of Manski (1997) with the monotone treatment selection (MTS) assumption of Manski and Pepper (2000) to obtain sharp bounds on the average treatment effect that are tighter than those obtained by imposing either set of assumptions alone.

In the context of discrete outcome data, Bhattacharya *et al.* (2008) use the Shaikh and Vytlacil (2005) strategy to bound the effect of Swan-Ganz catheterization on patient mortality, without imposing *a priori* knowledge of the sign in the MTR assumption. Boes (2009, 2010) derives sharp bounds on ordered outcome distributions by imposing a nonparametric threshold crossing model, and by taking advantage of a sequential information structure. Machado *et al.* (2009) consider the implications of weak monotonicity assumptions on the outcome and selection process in the binary/binary case, partly drawing on the results of the above literature, on Balke and Pearl (1997), and on Heckman and Vytlacil (2001). See also Chiburis

(2009) for a generalized linear model approach, and Chesher (2007, 2008) for a non-linear structural model approach.

A related strand of the literature concerns the empirical analysis of moment inequalities, as derived for example in Pakes *et al.* (2006) and Pakes (2009). Chernozhukov *et al.* (2007) and Andrews *et al.* (2004) establish general identification, estimation, and inference results for such models. See also Rosen (2008), Guggenberger *et al.* (2008), Andrews and Guggenberger (2009), Andrews and Han (2009), Romano and Shaikh (2009), and Stoye (2009). Most of this literature can be embedded within the broader topic of partially identified models (Manski 2003, Tamer 2009). Regarding discrete outcomes, Manski and Tamer (2002) discuss inference on regressions with interval data. Tamer (2003) analyzes the incompleteness of a simultaneous binary choice model. Inference is discussed, for example, in Imbens and Manski (2004).

The remainder of the paper is structured as follows. The next two sections propose several variations of the MIV and MTR assumptions. Section 4 investigates the identifying power when both sets of assumptions are imposed jointly. Section 5 illustrates the construction of bounds in a simulated data example and applies the proposed methods to bound the effect of education on smoking. Section 6 concludes.

2 Semi-Monotonicity in Subpopulations

2.1 Instrumental Variables Assumptions

Suppose that outcomes are (at least) measured on a discrete ordinal scale, i.e., responses can be ranked from low to high, but differences between outcomes are not necessarily defined. Each individual is assumed to have a response function, denoted by $Y(s) \in \mathcal{Y}$, that determines the outcome in each state $s \in \mathcal{S}$. Let $S \in \mathcal{S}$ denote the realized state, let $Y = Y(S)$ denote the realized outcome, and let $X = (W, V)$ denote the vector of covariates with support $\mathcal{X} = \mathcal{W} \times \mathcal{V}$. The observed data are the triple (Y, S, X) with distribution $P(Y, S, X)$.

The problem is to learn $P[Y(s) \leq y]$ or $P[Y(s) \leq y | W]$ for all $s \in \mathcal{S}$ from the observed data

distribution. Without imposing additional assumptions, point identification from $P(Y, S, X)$ fails because the outcome of each individual is only observed in the realized state, the outcomes that would have occured under alternative states are logically unobserved. This yields the no-assumptions bounds of Manski (1990):

$$P(Y \leq y|W, S = s) \cdot P(S = s|W) \ \leq \ P[Y(s) \leq y|W] \ \leq \tag{1}$$

$$P(Y \leq y|W, S = s) \cdot P(S = s|W) + P(S \neq s|W)$$

which follows by the law of total probability, and bounds zero and one on the counterfactual distributions $P(Y \leq y|W, S \neq s)$. The interval defined in (1) can also be interpreted as an identification region for the potential outcome distribution (Manski 2003).

Additional information can be gained using an instrumental variables strategy. Standard IV methods state that covariate V is an instrument if the distribution of outcomes for each $s \in \mathcal{S}$ conditional on W does not change with different values of V. Formally,

Assumption (IV). *Covariate V is an IV in the sense of conditional independence in the distribution of Y if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$,*

$$P[Y(s) \leq y|W, V = u_1] = P[Y(s) \leq y|W, V = u_2]. \tag{2}$$

Assumption IV implies that $P[Y(s) \leq y|W, V] = P[Y(s) \leq y|W]$ and thus a lower bound on the latter is given by the largest of the lower bounds on $P[Y(s) \leq y|W, V = u]$ over all u, and an upper bound is given by the smallest of the upper bounds on $P[Y(s) \leq y|W, V = u]$ over all u (Manski 1990, 1995).

Now assume that $\mathcal{V}$ is an ordered set. A distributional equivalent to the MIV assumption of Manski and Pepper (2000) would be that the distribution of $Y(s)$ for all individuals with a specified W and value u_1 of V weakly dominates the distribution of $Y(s)$ for individuals with the same W but with $u_2 \geq u_1$. Formally,

Assumption (MIV). *Let $\mathcal{V}$ be an ordered set. Covariate V is an IV in the sense of weak stochastic dominance in the distribution of Y if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all*

$u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ with $u_1 \leq u_2$,

$$P[Y(s) \leq y|W, V = u_1] \geq P[Y(s) \leq y|W, V = u_2]. \tag{3}$$

Depending on the application and the implications of the theoretical model, the sign of the inequality may be reversed such that the distribution of $Y(s)$ conditional on $V = u_2$ weakly dominates the distribution of $Y(s)$ conditional on $V = u_1$.

Assumption MIV implies for any $u \in \mathcal{V}$ with $u_1 \leq u \leq u_2$ and $\forall y \in \mathcal{Y}$ that

$$P[Y(s) \leq y|W, V = u_1] \quad \geq \quad P[Y(s) \leq y|W, V = u] \quad \geq \quad P[Y(s) \leq y|W, V = u_2]. \tag{4}$$

Thus, a lower bound of $P[Y(s) \leq y|W, V = u_2]$ is also a lower bound of $P[Y(s) \leq y|W, V = u]$, and an upper bound of $P[Y(s) \leq y|W, V = u_1]$ is also an upper bound of $P[Y(s) \leq y|W, V = u]$. This must hold for all values $u_1 \leq u$ and must hold for all values $u_2 \geq u$. Exploiting the full support of V, the following sharp bounds on the distribution of potential outcomes can be derived (see Manski and Pepper 2000, but also Blundell *et al.* 2007):

$$\inf_{u_1 \leq u} \left\{ P[Y \leq y|W, S = s, V = u_1] \cdot P(S = s|W, V = u_1) + P(S \neq s|W, V = u_1) \right\}$$

$$\geq \quad P[Y(s) \leq y|W, V = u] \quad \geq \tag{5}$$

$$\sup_{u_2 \geq u} \left\{ P[Y \leq y|W, S = s, V = u_2] \cdot P(S = s|W, V = u_2) \right\}.$$

Bounds on $P[Y(s) \leq y|W]$ can be obtained by integrating out over the support of V (subject to measurability conditions). Two remarks can be made regarding the properties of the bounds in (5). First, they are not informative if the lower and upper no-assumptions bounds on $P[Y(s) \leq y|W, V = u]$ weakly decrease with u, in which case they coincide with the no-assumptions bounds. Second, the bounds coincide with the IV bounds if the no-assumptions lower and upper bounds on $P[Y(s) = y|W, V = u]$ weakly increase with u.

Assumption MIV might be too restrictive if monotonicity only holds for subsets of $\mathcal{V}$. For example, one might think of a combination of monotonically increasing and monotonically decreasing distribution functions $P[Y(s) \leq y|W, V = u]$ over the support of V. Formally,

6

Assumption (MMIV). *Let $\mathcal{V}$ be an ordered set. Covariate V is an IV in the sense of partial weak stochastic dominance in the distribution of Y with a global maximum if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ with $u_1 \leq u_2 \leq u_{\max} = \arg\max_{u \in \mathcal{V}} P[Y(s) \leq y | W, V = u]$,*

$$P[Y(s) \leq y | W, V = u_1] \ \leq \ P[Y(s) \leq y | W, V = u_2], \tag{6}$$

and for all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ with $u_{\max} \leq u_1 \leq u_2$,

$$P[Y(s) \leq y | W, V = u_1] \ \geq \ P[Y(s) \leq y | W, V = u_2]. \tag{7}$$

Assumption MMIV (the maximum monotone instrumental variables assumption) yields the assumption MIV as a special case if the global maximum point is at the lower boundary of $\mathcal{V}$. While assumption MMIV allows for a single break in the monotonicity of $P[Y(s) \leq y | W, V = u]$ over the support of V, multiple break points would be possible as well. However, the identification of a global optimum is generally easier to establish from economic theory than multiple local optima. Depending on the application, assumption MMIV might be reversed to a minimum condition (justifying the more general semi-monotonicity terminology).

In order to analyze the implications of assumption MMIV on the (partial) identification of $P[Y(s) \leq y | W, V = u]$, define the following subsets of $\mathcal{V}$:

$$\mathcal{V}_{\max}^l \ = \ \{u \in \mathcal{V} : u \leq \arg\max_{u \in \mathcal{V}} P[Y(s) \leq y | W, V = u]\}$$

$$\mathcal{V}_{\max}^u \ = \ \{u \in \mathcal{V} : u \geq \arg\max_{u \in \mathcal{V}} P[Y(s) \leq y | W, V = u]\}.$$

For any $u \in \mathcal{V}_{\max}^u$ with $u_1 \leq u \leq u_2$ and evaluation points $u_1, u_2 \in \mathcal{V}_{\max}^u \times \mathcal{V}_{\max}^u$, and $\forall y \in \mathcal{Y}$, assumption MMIV implies the same set of inequalities as in (4), and thus the same bounds on $P[Y(s) \leq y | W, V = u]$. For any $u \in \mathcal{V}_{\max}^l$ with $u_1 \leq u \leq u_2$ and $u_1, u_2 \in \mathcal{V}_{\max}^l \times \mathcal{V}_{\max}^l$, and $\forall y \in \mathcal{Y}$, assumption MMIV implies

$$P[Y(s) \leq y | W, V = u_1] \ \leq \ P[Y(s) \leq y | W, V = u] \ \leq \ P[Y(s) \leq y | W, V = u_2]. \tag{8}$$

Thus, an upper bound of $P[Y(s) \leq y | W, V = u_2]$ is also an upper bound of $P[Y(s) \leq y | W, V = u]$, and a lower bound of $P[Y(s) \leq y | W, V = u_1]$ is also a lower bound of $P[Y(s) \leq y | W, V = u]$.

This must hold for all values $u_1 \leq u$ and for all values $u_2 \geq u$ as defined above. Exploiting the full support of V yields the following sharp bounds:

Proposition 1. *Let assumption* MMIV *hold. Then for any* $u \in \mathcal{V}_{\max}^l$,

$$
\sup_{u_1 \leq u \in \mathcal{V}_{\max}^l} \left\{ P[Y \leq y | W, S = s, V = u_1] \cdot P(S = s | W, V = u_1) \right\}
$$
$$
\leq \quad P[Y(s) \leq y | W, V = u] \quad \leq \tag{9}
$$
$$
\inf_{u_2 \geq u \in \mathcal{V}_{\max}^l} \left\{ P[Y \leq y | W, S = s, V = u_2] \cdot P(S = s | W, V = u_2) + P(S \neq s | W, V = u_2) \right\}.
$$

and for any $u \in \mathcal{V}_{\max}^u$,

$$
\sup_{u_2 \geq u \in \mathcal{V}_{\max}^u} \left\{ P[Y \leq y | W, S = s, V = u_2] \cdot P(S = s | W, V = u_2) \right\}
$$
$$
\leq \quad P[Y(s) \leq y | W, V = u] \quad \leq \tag{10}
$$
$$
\inf_{u_1 \leq u \in \mathcal{V}_{\max}^u} \left\{ P[Y \leq y | W, S = s, V = u_1] \cdot P(S = s | W, V = u_1) + P(S \neq s | W, V = u_1) \right\}.
$$

In the absence of other information, these bounds are sharp.

Compared to assumption MIV, the construction of bounds under MMIV requires a split of the support of V into two parts, one for the lower level set and one for the upper level set relative to the global maximum point (including the maximum as a boundary). These two sets yield the bounds in (9) and (10), respectively. Note that weak monotonicity in the no-assumptions lower and upper bounds on $P[Y(s) \leq y | W, V = u]$ implies that either (9) or (10) coincides with the no-assumptions bounds or the bounds under IV, and *vice versa*. Note too that Proposition 1 requires credible knowledge about $u_{\max}$.

In order to motivate the foregoing assumptions, consider the following example. Suppose one is interested in the effect of education (S) on individual health (Y). Assume that education outcomes are generated according to an ordered choice model with the optimal stopping time determined by pairwise comparisons of returns in adjacent schooling stages. A common assumption in this model is the concavity of the marginal return function; see Cunha *et al.* (2008) for a discussion of this type of model and extensions. Let the threshold values depend on transition-specific characteristics (such as school subsidies or tuition fees), some of which

are unobserved, and assume, for simplicity, that there are three levels of subsidies: no subsidy ($V = u_1$), modest subsidy ($V = u_2$), and maximum subsidy ($V = u_3$).

For V to serve as an instrument in the sense of IV we would assume that the potential health distribution does not change with the level of the subsidy. This assumption might be reasonable if subsidies are assumed to be independent of health behavior. If, however, the smarter people are more likely to get a grant, and they are also better able to take care of themselves, then assumption IV would be invalid (support for the existence of such a channel is provided, for example, by Cutler and Lleras-Muney, 2009). However, assumption MIV would be reasonable if transition-specific subsidies increase the education outcomes and are symmetrically related to health behavior (generating a weak stochastic dominance with increasing V; the lower V, the lower the potential health).

If there is an asymmetry in the link to individual health, then monotonicity might be combined with a weak concavity assumption. For instance, if education outcomes are generally higher the higher the subsidies, and if the transition to the higher education levels are more relevant for generating positive health (Cutler and Lleras-Muney, 2009, for related arguments), then the dominance of the distribution of potential health conditional on $V = u_1$ relative to $V = u_2$ might be smaller than that of $V = u_2$ relative to $V = u_3$. In the extreme case, this might yield a maximum point at $V = u_2$, and thus conform with MMIV. The data examples below will further illustrate this point.

2.2 Semi-Monotone Treatment Selection

A special case of assumptions IV, MIV, and MMIV is obtained if V is the treatment variable itself. In this case, the IV assumption is equivalent to the exogenous treatment selection assumption, stating that the potential outcome distribution does not vary with S, i.e., for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and $u_1, u_2 \in \mathcal{S} \times \mathcal{S}$,

$$P[Y(s) \leq y | W, S = u_1] = P(Y(s) \leq y | W, S = u_2]. \tag{11}$$

This assumption point identifies the counterfactual distribution since $P[Y(s) \leq y|W, S \neq s] = P[Y \leq y|W, S = s]$, and thus $P[Y(s) \leq y|W] = P(Y \leq y|W, S = s)$.

The MIV assumption becomes a monotone treatment selection (MTS) condition asserting that the potential outcome distribution given $S = u_1$ weakly dominates the distribution conditional on $S = u_2$ with $u_1 \leq u_2$. The MTS assumption implies that

$$
\begin{aligned}
u < s &\Rightarrow P(Y \leq y|W, S = s) \leq P[Y(s) \leq y|W, S = u] \leq 1 \\
u = s &\Rightarrow P[Y(s) \leq y|W, S = u] = P(Y \leq y|W, S = s) \\
u > s &\Rightarrow 0 \leq P[Y(s) \leq y|W, S = u] \leq P(Y \leq y|W, S = s),
\end{aligned}
\tag{12}
$$

and thus

$$
\begin{aligned}
P[Y \leq y|W, S = s] \cdot P(S \leq s|W) &\leq P[Y(s) \leq y|W] \leq \\
P[Y \leq y|W, S = s] \cdot P(S \geq s|W) &+ P(S < s|W).
\end{aligned}
\tag{13}
$$

The MMIV assumption with $V = S$ postulates the existence of a unique global maximum point $u_{\max} = \arg\max_{u \in \mathcal{S}} P[Y(s) \leq y|W, S = u]$. The potential outcome dstribution is assumed to be monotonically increasing for all states $u \leq u_{\max}$ and monotonically decreasing for all states $u \geq u_{\max}$. Corollary 1 summarizes the bounds implied by this maximum monotone treatment selection (MMTS) assumption.

Corollary 1. *Let assumption MMIV hold with $V = S$. Then*

$$
\begin{aligned}
\left.\begin{array}{l} s < u \leq u_{\max} \\ u_{\max} \leq u < s \end{array}\right\} &\Rightarrow P(Y \leq y|W, S = s) \leq P[Y(s) \leq y|W, S = u] \leq 1 \\
u = s &\Rightarrow P[Y(s) \leq y|W, S = u] = P(Y \leq y|W, S = s) \\
\left.\begin{array}{l} u < s \leq u_{\max} \\ u_{\max} \leq s < u \end{array}\right\} &\Rightarrow 0 \leq P[Y(s) \leq y|W, S = u] \leq P(Y \leq y|W, S = s) \\
else &\Rightarrow 0 \leq P[Y(s) \leq y|W, S = u] \leq 1,
\end{aligned}
$$

and thus

$$
s < u_{\max} \Rightarrow P(Y \leq y|W, S = s) \cdot P(s \leq S \leq u_{\max}|W)
\tag{14}
$$

$$\leq\ P[Y(s) \leq y|W]\ \leq\ P(Y \leq y|W, S = s) \cdot P(S \leq s|W) + P(S > s|W)$$

$$s = u_{\max}\ \Rightarrow\ P(Y \leq y|W, S = s) \cdot P(S = s|W) \tag{15}$$

$$\leq\ P[Y(s) \leq y|W]\ \leq\ P(Y \leq y|W, S = s)$$

$$s > u_{\max}\ \Rightarrow\ P(Y \leq y|W, S = s) \cdot P(u_{\max} \leq S \leq s|W) \tag{16}$$

$$\leq\ P[Y(s) \leq y|W]\ \leq\ P(Y \leq y|W, S = s) \cdot P(S \geq s|W) + P(S < s|W).$$

In the absence of other information, these bounds are sharp.

The bounds in Corollary 1 result from a split of $\mathcal{S}$ into two subsets with separation point equal to the global maximum $u_{\max}$. If s and u both lie on one side of $u_{\max}$, then the MMTS assumption is informative and yields bounds that are tighter than the logical unit range. If s and u are not element of the same subset, then the MMTS assumption alone is not informative regarding the counterfactual distribution.

3 Semi-Monotone Treatment Response

The preceding section explored assumptions in terms of an instrument that, when conditioned on, restricts the shape of the potential outcome distribution. One might also think of imposing assumptions on the response function $Y(s)$ directly, either for each individual, or for the distribution of outcomes in the population. Assumptions of the former type have been investigated in Manski (1997), Manski and Pepper (2000, 2009), and Okumura and Usui (2007). Two assumptions of the latter type, referred to as weak monotone treatment response (WMTR) and maximum monotone treatment response (MMTR), will be investigated here:

Assumption (WMTR). *Let $\mathcal{S}$ be an ordered set. $Y(s)$ fulfills a weak stochastic order condition if, for each $y \in \mathcal{Y}$, $t \in \mathcal{S}$, $X \in \mathcal{X}$, and all $s_1, s_2 \in \mathcal{S} \times \mathcal{S}$ with $s_1 \leq s_2$,*

$$P[Y(s_1) \leq y|X, S = t]\ \geq\ P[Y(s_2) \leq y|X, S = t]. \tag{17}$$

Assumption (MMTR). *Let $\mathcal{S}$ be an ordered set. $Y(s)$ fulfills a weak stochastic order condition with a maximum point if, for each $y \in \mathcal{Y}$, $t \in \mathcal{S}$, $X \in \mathcal{X}$, and all $s_1, s_2 \in \mathcal{S} \times \mathcal{S}$ with*

$s_1 \leq s_2 \leq s_{\max} = \arg\max_{s \in \mathcal{S}} P[Y(s) \leq y | X, S = t]$,

$$P[Y(s_1) \leq y | X, S = t] \ \leq \ P[Y(s_2) \leq y | X, S = t], \tag{18}$$

and for all $s_1, s_2 \in \mathcal{S} \times \mathcal{S}$ with $s_{\max} \leq s_1 \leq s_2$,

$$P[Y(s_1) \leq y | X, S = t] \ \geq \ P[Y(s_2) \leq y | X, S = t]. \tag{19}$$

Assumption WMTR states that the outcome distribution in state s_1 weakly dominates the outcome distribution in state $s_2 \geq s_1$, irrespective of the realized state. WMTR holds, for example, under the monotone treatment response assumption of Manski (1997) that restricts individual responses by the condition: $s_2 \geq s_1 \Rightarrow Y(s_2) \geq Y(s_1)$. Assumption WMTR is also fulfilled if the Manski (1997) assumption does not hold for all members of the population, but on average across the population. The weak inequality is assumed to hold conditional on X, and thus still holds conditional on W, or unconditional on covariates.

Assumption MMTR asserts the existence of a state $s_{\max}$ that stochastically dominates all other states, i.e., $P[Y(s_{\max}) \leq y | X, S = t] \geq P[Y(s) \leq y | X, S = t)$ for all $s \neq s_{\max}$. The distributions are assumed to be monotonically increasing in s to the left and monotonically decreasing to the right of the maximum point. Hence, the MMTR assumption encompasses the WMTR assumption as a special case if $s_{\max}$ is at the lower boundary of $\mathcal{S}$. It may also reverse the sign of the weak inequality if the maximum point is at the upper boundary of $\mathcal{S}$. It is straightforward to replace MMTR by a minimum condition, so one may refer to it more generally as a semi-monotone treatment response assumption.

Assumption WMTR places the following sharp bounds on the potential outcome distribution (see Manski 1997 for further details):

$$P(Y \leq y | W, S \geq s) P(S \geq s | W) \ \leq \ P[Y(s) \leq y | W] \ \leq \tag{20}$$

$$P(Y \leq y | W, S \leq s) P(S \leq s | W) + P(S > s | W).$$

which follows from the law of total probability and the presumed monotonicity, such that $P(Y \leq y | W, S = s)$ can be used as an informative lower bound on all $P[Y(s) \leq y | W, S = t)$ with $t > s$, and as an informative upper bound on all $P[Y(s) \leq y | W, S = t)$ with $t < s$.

Weak monotonicity combined with a global maximum point splits the support of $\mathcal{S}$ into a lower level and an upper level set relative to $s_{\max}$. While the same type of bounds on the counterfactual probabilities as under WMTR apply for the latter, arguments with a reversed sign of the inequality can be applied for the former. Proposition 2 summarizes the bounds:

Proposition 2. *Let assumption MMTR hold. Then*

$$\left. \begin{array}{l} s < t \leq s_{\max} \\ s_{\max} \leq t < s \end{array} \right\} \quad \Rightarrow \quad 0 \;\leq\; P[Y(s) \leq y | W, S = t] \;\leq\; P(Y \leq y | W, S = t)$$

$$s = t \quad \Rightarrow \quad P[Y(s) \leq y | W, S = t] \;=\; P(Y \leq y | W, S = s)$$

$$\left. \begin{array}{l} t < s \leq s_{\max} \\ s_{\max} \leq s < t \end{array} \right\} \quad \Rightarrow \quad P(Y \leq y | W, S = t) \;\leq\; P[Y(s) \leq y | W, S = t] \;\leq\; 1$$

$$else \quad \Rightarrow \quad 0 \;\leq\; P[Y(s) \leq y | W, S = t] \;\leq\; 1,$$

and thus

$$s < s_{\max} \quad \Rightarrow \quad P(Y \leq y | W, S \leq s) \cdot P(S \leq s | W) \;\leq\; P[Y(s) \leq y | W] \;\leq\; \tag{21}$$

$$P(Y \leq y | W, s \leq S \leq s_{\max}) \cdot P(s \leq S \leq s_{\max} | W)$$

$$+[1 - P(s \leq S \leq s_{\max} | W)]$$

$$s = s_{\max} \quad \Rightarrow \quad P(Y \leq y | W) \;\leq\; P[Y(s) \leq y | W] \;\leq\; \tag{22}$$

$$P(Y \leq y | W, S = s) \cdot P(S = s | W) + [1 - P(S = s | W)]$$

$$s > s_{\max} \quad \Rightarrow \quad P(Y \leq y | W, S \geq s) \cdot P(S \geq s | W) \;\leq\; P[Y(s) \leq y | W] \;\leq\; \tag{23}$$

$$P(Y \leq y | W, s_{\max} \leq S \leq s) \cdot P(s_{\max} \leq S \leq s | W)$$

$$+[1 - P(s_{\max} \leq S \leq s | W)].$$

In the absence of other information, these bounds are sharp.

Note that assumptions MMTR and MMTS contribute differently to the identification of $P[Y(s) \leq y | W]$. While the former explores stochastic dominance over potential states s in $Y(s)$, the latter explores stochastic dominance over realized states, i.e., conditional on S. In each case, credible knowledge of the stochastic order, i.e., the sign of the weak inequalities, and of the global maximum points is essential in the construction of bounds.

13

4 Assumptions on Treatment Selection and Response

One might also think of combining the assumptions of Sections 2 and 3. A particularly interesting case is obtained if the MMTS and the MMTR assumptions are imposed jointly. In order to illustrate their identifying power, it is helpful to employ a matrix representation of the counterfactual probabilities, as in the following example with four states:

$$
\begin{array}{ccccccc}
P[Y(1) \leq y|W, S = 1] & \leq & P[Y(1) \leq y|W, S = 2] & \leq & P[Y(1) \leq y|W, S = 3] & \geq & P[Y(1) \leq y|W, S = 4] \\
\leq & & \leq & & \leq & & \leq \\
P[Y(2) \leq y|W, S = 1] & \leq & P[Y(2) \leq y|W, S = 2] & \leq & P[Y(2) \leq y|W, S = 3] & \geq & P[Y(2) \leq y|W, S = 4] \\
\leq & & \leq & & \leq & & \leq \\
P[Y(3) \leq y|W, S = 1] & \leq & P[Y(3) \leq y|W, S = 2] & \leq & P[Y(3) \leq y|W, S = 3] & \geq & P[Y(3) \leq y|W, S = 4] \\
\geq & & \geq & & \geq & & \geq \\
P[Y(4) \leq y|W, S = 1] & \leq & P[Y(4) \leq y|W, S = 2] & \leq & P[Y(4) \leq y|W, S = 3] & \geq & P[Y(4) \leq y|W, S = 4]
\end{array}
$$

The main diagonal elements are identified from the observed data distribution, the off-diagonal elements are not identified. The MMTS assumption places a sequence of inequalities between the columns of the matrix, the MMTR assumption places inequalities between the rows of the matrix. Suppose, for simplicity, that $s_{\max} = u_{\max} = 3$ such that the inequalities can be placed as shown above. The matrix can be used to derive sharp lower and upper bounds on $P[Y(s) \leq y||W, S = u]$ for all $s \neq u$. For example, $P[Y(2) \leq y|W, S = 1]$ can be bounded from below by $P(Y \leq y|W, S = 1)$ and bounded from above by $P(Y \leq y|W, S = 2)$. Informative bounds can be placed on all counterfactual distributions with $s \leq s_{\max}$ and $u \leq u_{\max}$, or $s \geq s_{\max}$ and $u \geq u_{\max}$. For all other s, u combinations, the bounds zero and one apply.

The bounds on the potential outcome distributions $P[Y(s) \leq y|W]$ in the special case of MMTS and MMTR with $s_{\max} = u_{\max}$ are summarized in the following corollary:

Corollary 2. *Let assumption MMIV hold with $V = S$, and let assumption MMTR hold. Suppose $s_{\max} = u_{\max}$. Then,*

$$
s < s_{\max} \quad \Rightarrow \quad P(Y \leq y|W, S \leq s) \cdot P(S \leq s|W) + P(Y \leq y|W, S = s) \tag{24}
$$

$$
\cdot P(s < S \leq s_{\max}|W) \ \leq \ P[Y(s) \leq y|W] \ \leq
$$

$$
P(Y \leq y|W, S = s) \cdot P(S < s|W) + P(Y \leq y|W, s \leq S \leq s_{\max})
$$

14

$$\cdot P(s \leq S \leq s_{\max}|W) + P(S > s_{\max}|W),$$

$$s = s_{\max} \quad \Rightarrow \quad P(Y \leq y|W) \ \leq \ P[Y(s) \leq y|W] \ \leq \ P(Y \leq y|W, S = s), \tag{25}$$

$$s > s_{\max} \quad \Rightarrow \quad P(Y \leq y|W, S \geq s) \cdot P(S \geq s|W) + P(Y \leq y|W, S = s) \tag{26}$$

$$\cdot P(s_{\max} \leq S < s|W) \ \leq \ P[Y(s) \leq y|W] \ \leq$$

$$P(S < s_{\max}|W) + P(Y \leq y|W, s_{\max} \leq S \leq s) \cdot P(s_{\max} \leq S \leq s|W)$$

$$+ P(Y \leq y|W, S = s) \cdot P(S > s|W)].$$

In the absence of other information, these bounds are sharp.

The key to Corollary 2 is again to split the support of S into a lower and an upper level set relative to the global maximum point, and to derive the bounds on the counterfactual probabilities for each of these sets. Note that the bounds of Corollary 2 are never wider but often tighter than the bounds of Corollary 1 or the bounds of Proposition 2.

5 Illustrations

5.1 Simulated Data Example

Consider the following model for the selection and outcome processes:

$$S = \sum_{s=1}^{3} s\mathcal{I}(\tau_{s-1} < \psi \leq \tau_s) \tag{27}$$

$$Y(s) = \sum_{y=1}^{3} y\mathcal{I}(\kappa_{s,y-1} < \varepsilon \leq \kappa_{s,y}), \quad s = 1, 2, 3$$

where $\mathcal{I}(A)$ is the logical indicator function that returns one when A is true. The observed data are the pair (Y, S) with $Y = Y(S)$. The selection status and the potential outcomes in each state are generated according to a threshold crossing mechanism where $-\infty = \tau_0 < \tau_1 < \tau_2 < \tau_3 = \infty$, and $-\infty = \kappa_{s,0} < \kappa_{s,1} < \kappa_{s,2} < \kappa_{s,3} = \infty$. The selection model allows for random thresholds with $\tau_1 = \nu_1$ and $\tau_2 = \tau_1 + \exp(\nu_2)$ to fulfill the order condition. The parameters $\kappa_{s,y}$ and ψ are constant but unknown.

Model (27) does not impose *a-priori* restrictions on the joint distribution of $(\nu_1, \nu_2, \varepsilon)$. In particular, the correlations between the error terms of the selection process and the error term of the outcome process are free to vary (except for common regularity conditions such as a finite and positive definite covariance matrix). Moreover, the model allows for arbitrary behavioral assumptions in the selection process since no *a-priori* assumptions are imposed on the correlation between ν_1 and ν_2. The model does however impose the restriction that the error terms in all states are generated by the same process. This assumption substantially simplifies the illustration below, by allowing to focus on the influence of the main parameters on the construction and properties of the bounds.

The model can be extended to allow for covariates W by making the thresholds in the selection process and in the outcome processes functions of W. However, in line with the notion of the paper, the model avoids exclusion restrictions (i.e, non-overlapping subsets of W appear in κ and τ), such that accounting for covariates amounts to defining treatment effects (and deriving bounds) for subpopulations described by W.

In order to illustrate the construction of bounds, the error terms $(\nu_1, \nu_2, \varepsilon)$ are assumed jointly normal with zero means and covariance matrix

$$
\Sigma = \begin{pmatrix} 1 & \sigma_{1,2} & \sigma_{1,\varepsilon} \\ \sigma_{1,2} & 1 & \sigma_{2,\varepsilon} \\ \sigma_{1,\varepsilon} & \sigma_{2,\varepsilon} & 1 \end{pmatrix}.
$$

The following additional restrictions are imposed on the data-generating process: $\sigma_{1,\varepsilon} = \alpha/2$, $\sigma_{1,\varepsilon} = -\alpha/2$, $\sigma_{1,2} = 0$. The thresholds in the outcome processes are specified as

$$
\kappa_{1,1} = -\beta/2 \quad \kappa_{2,1} = 0 \quad \kappa_{3,1} = -\beta/5 \quad \text{(lower thresholds)}
$$

$$
\kappa_{1,2} = .7 \qquad \kappa_{2,2} = .3 \quad \kappa_{3,2} = .4 \qquad \text{(upper thresholds)}
$$

The parameters α (scenario 1), β (scenario 2), and ψ (scenario 3) are varied in the simulation study, one at a time between -1 and 1 with the other parameters fixed at one.

Each of these parameters makes a different contribution to the (partial) identification of the potential outcome distribution. The parameter α does only affect the treatment selection

16

assumption. For α positive, the assumption of an asymmetric correlation structure (positive correlation for the lower selection threshold and negative correlation for the upper selection threshold) generates a global maximum point in the MMTS sense at $u_{\max} = 2$. If α is negative, then the maximum point turns into a global minimum point. If α equals zero, then the potential outcome distribution does not vary conditional on different values of S, which is equivalent to an IV assumption with the instrument equal to the treatment variable.

The parameter β only affects the treatment response assumption because it changes the lower outcome thresholds. If β is positive, then $P[Y(s) \leq 1|S]$ is maximal at $s_{\max} = 2$, which turns into a minimum point if β is negative. If β equals zero, then $P[Y(s) \leq 1|S]$ does not differ over states. The upper outcome thresholds are chosen such that the order condition is fulfilled, and the relative magnitude over states generates a global minimum point in $P[Y(s) \leq 2|S]$ at $s_{\min} = 2$. Finally, ψ does not affect the treatment selection and treatment response assumptions, but it affects the severity of the partial identification result, i.e., the width of the bounds, by varying the proportion of observations in each state.

Figures 1 to 3 display the potential outcome distribution if α is varied from -1 to 1 (and all other parameters are fixed as described above). In each figure, the diagrams on the left show the probability that the potential outcome takes value 1 in state 1 (top), in state 2 (middle), and in state 3 (bottom). The right diagrams show the probability that the potential outcome takes values less or equal 2 in the three states. Figure 1 shows the bounds on the potential outcome distribution if only the MMTS assumption is imposed, Figure 2 if only the MMTR assumption is imposed, and Figure 3 if both assumptions are imposed jointly.

— Insert Figures 1 to 3 about here —

Consider Figure 1 and the top left diagram, i.e., the probability that the potential outcome in state 1 takes value 1. This probability can be written as

$$
\begin{aligned}
P[Y(1) \leq 1] = {} & P(Y \leq 1|S = 1)P(S = 1) + P[Y(1) \leq 1|S = 2]P(S = 2) \\
& + P[Y(1) \leq 1|S = 3]P(S = 3)
\end{aligned}
$$

where all terms but $P[Y(1) \leq 1|S = 2]$ and $P[Y(1) \leq 1|S = 3]$ are point identified from the observed data distribution $P(Y, S)$. If α is negative, then the MMTS assumption implies a global minimum point at $2 = \arg\min_{u \in \mathcal{S}} P[Y(s) \leq 1|S = u]$ such that $P(Y \leq 1|S = 1)$ can be used as an upper bound on $P[Y(1) \leq 1|S = 2]$ instead of one with the lower bound zero unchanged. $P[Y(1) \leq 1|S = 3]$ is bounded by zero and one since it lies to the right of the minimum point with the identified probability lying on the left and thus no informative conclusions can be drawn from the MMTS assumption. If α is positive, then the minimum point is replaced by a maximum point such that $P(Y \leq 1|S = 1)$ can be used as a lower bound on $P[Y(1) \leq 1|S = 2]$ instead of zero with the no-assumptions upper bound one unchanged. If $\alpha = 0$, then the potential outcome distribution does not vary conditional on different S. and thus $P[Y(1) \leq 1]$ is point identified by the observed $P(Y \leq 1|S = 1)$. Similar arguments apply to the cumulative probabilities of outcome 2 and the other states.

Now consider the top left diagram of Figure 2. The assumptions imposed on the data-generating process imply a global maximum point in the MMTR sense at $2 = \arg\max_{s \in \mathcal{S}} P[Y(s) \leq 1|S = u]$. Thus, irrespective of α, $P(Y \leq 2|S = 2)$ can be used as an upper bound for $P[Y(1) \leq 1|S = 2]$ instead of one with the no-assumptions lower bound zero unchanged, but no informative conclusions can be drawn from the MMTR assumption on $P[Y(1) \leq 1|S = 3]$. By the assumptions on the upper thresholds of the outcome process, the global maximum point turns into a global minimum point for outcomes less or equal two, and thus $P(Y \leq 2|S = 2)$ can be used as an informative lower bound on $P[Y(1) \leq 2|S = 2]$ with the no-assumption upper bound one unchanged. Again, similar arguments apply to the other states.

Figure 3 combines the bounds of Figures 1 and 2. If the MMTS and MMTR assumptions are imposed jointly, then the larger of the lower bounds and the smaller of the upper bounds under each assumption provides a binding constraint on the possible range of the potential outcome distribution. Since the two assumptions make different contributions on identification, the overlap of both assumptions may provide a substantial improvement over the bounds obtained under either assumption alone. This explains, on the one hand, the observed kinks in the

upper and lower bounds, and, on the other hand, the small width in some cases (such as for the lower part of the distribution in state 2 and $\alpha > 0$, and the upper part for $\alpha < 0$.

Figures 4 to 6 summarize the results for β altered from -1 to 1 (scenario 2). Since β does not affect the MMTS assumption, and α is fixed at 1, the conclusions drawn from the MMTS assumption are analogous to the outer right part of the bounds in Figure 1. Furthermore, since β only affects the lower threshold parameters, a change in the bounds over the range of β is only observed in the left panel of Figure 5. If β is negative, then $2 = \arg\min_{s \in \mathcal{S}} P[Y(s) \leq 1 | S = u]$ such that, for example, $P(Y \leq 1 | S = 2)$ can be used as a lower bound on $P[Y(1) \leq 1 | S = 2]$ instead of zero, with the no-assumptions upper bound one unchanged. If β is positive, then the minimum point turns into a maximum point such that $P(Y \leq 1 | S = 2)$ can be used as an upper bound on $P[Y(1) \leq 1 | S = 2]$ instead of one, with the no-assumptions lower bound zero unchanged. This explains the upward shift in the lower bound for $\beta < 0$ and the downward shift in the upper bound for $\beta > 0$. If $\beta = 0$, the potential outcome distribution is point identified because the thresholds do not vary over treatment states.

— Insert Figures 4 to 6 about here —

Scenario 3 considers the variation in ψ from -1 to 1. Since ψ only affects the distribution of states, but does not affect MMTS and MMTR (which are determined by $\alpha = \beta = 1$), this scenario illustrates the trade-off in the collection of data on states. The smaller ψ, the more observations are in state 1; the higher ψ, the more observations are in state 3. Thus, for low values of ψ, the bounds on the potential outcome distribution in state 1 is tightest, but the bounds on the distribution in state 3 are wide, and *vice versa* for high values of ψ. An (inverse) u-shaped effect of ψ on the bounds can be observed for the distribution in state 2. Again, if both assumptions are imposed jointly, then MMTS and MMTR may yield a substantial improvement over the no-assumptions bounds.

— Insert Figures 7 to 9 about here —

5.2 The Effect of Education on Smoking

As an empirical example, I consider the effect of schooling (S) on health behavior (Y), the latter measured as the number of cigarette packs smoked per day. The relationship between education and smoking has been thoroughly studied before (e.g., Grossman 1972, Farrell and Fuchs 1982, Kenkel 1991, Rosenzweig 1995, Currie and Moretti 2003, Kenkel *et al.* 2006, de Walque 2007, Grimard and Parent 2007, Gilman *et al.* 2008, and Tenn *et al.* 2008). The evidence suggests that the more educated people have a smaller likelihood to smoke, a higher likelihood to quit, and smoke less on average. The literature is not conclusive, however, regarding the magnitude of the impact, ranging from negligibly small to negative and significant. The studies differ in the identification strategies and data used to estimate the causal effect, if present, though none of them has explicitly accounted for schooling effects on the entire smoking distribution. The framework above is well-suited to address this problem.

The analysis is based on the Smoking Supplement of the 1979 US National Health Interview Survey which contains information on the respondent's socioeconomic characteristics and smoking behavior. Details on the data can be found in Mullahy (1985). The population has been restricted to employed white men aged between 35 and 60, which gives a total of 1,981 observations. Table 1 provides estimates of the smoking distribution conditional on schooling. The results indicate a global minimum in $P(Y \leq y|S)$ between 9 and 11 years of schooling. Thus, the joint (modified) assumptions MMTS with $2 = \arg\min_{u \in \mathcal{S}} P[Y(s) \leq y|S = u]$ and MMTR with $2 = \arg\min_{s \in \mathcal{S}} P[Y(s) \leq y|S = u]$ are not rejected by the empirical evidence alone, and I will maintain these assumptions in what follows.

— Insert Table 1 about here —

Figure 10 displays the bounds on the potential smoking distributions for each of the stated schooling levels. The no-assumptions bounds provide the widest consensus possible, imposing bounds zero and one on each counterfactual quantity (left bars). The bounds obtained under the MMTS assumption (middle left bars) make use of the monotonicity condition to the left

20

and to the right of the global minimum point. Similar arguments apply for the bounds under the MMTR assumption (middle right bars). The matrix representation provides an intuitive way to determine which bounds apply:

$$
\begin{array}{ccccccccc}
P[Y|S=1] & \geq & P[Y(1)|S=2] & \leq & P[Y(1)|S=3] & \leq & P[Y(1)|S=4] & \leq & P[Y(1)|S=5] \\
\geq & & \geq & & \geq & & \geq & & \geq \\
P[Y(2)|S=1] & \geq & P[Y|S=2] & \leq & P[Y(2)|S=3] & \leq & P[Y(2)|S=4] & \leq & P[Y(2)|S=5] \\
\leq & & \leq & & \leq & & \leq & & \leq \\
P[Y(3)|S=1] & \geq & P[Y(3)|S=2] & \leq & P[Y|S=3] & \leq & P[Y(3)|S=4] & \leq & P[Y(3)|S=5] \\
\leq & & \leq & & \leq & & \leq & & \leq \\
P[Y(4)|S=1] & \geq & P[Y(4)|S=2] & \leq & P[Y(4)|S=3] & \leq & P[Y|S=4] & \leq & P[Y(4)|S=5] \\
\leq & & \leq & & \leq & & \leq & & \leq \\
P[Y(5)|S=1] & \geq & P[Y(5)|S=2] & \leq & P[Y(5)|S=3] & \leq & P[Y(5)|S=4] & \leq & P[Y|S=5]
\end{array}
$$

where the weak inequalities between columns follow from the MMTS assumption, and the weak inequalities between rows follow from the MMTR assumption. Imposing both assumptions jointly yields an intersection of the bounds on each counterfactual distribution, and thus the overall bounds on the potential outcome distribution (right bars) are never wider than the intersection of the bounds under each assumption alone. Moreover, the bounds can be tighter than the intersection of bounds since MMTS and MMTR both can make contributions to more informative lower and more informative upper bounds.

— Insert Figure 10 about here —

The bounds in Figure 10 demonstrate that the MMTS and the MMTR assumptions jointly can have substantial identifying power compared to the no-assumptions bounds. For example, the possible range of the probability of non-smoking if everybody in the population were to receive 12 years of schooling shrinks from $[0.201; 0.857]$ to $[0.518; 0.674]$. However, the bounds are still too wide to identify the stochastic order of the distributions since they share common elements over all schooling levels. Thus, the analysis presented here cannot reject the hypothesis that the statistical association between schooling and smoking is only caused by unobserved confounding factors.

6 Conclusion

Constraints on functional form provide a powerful yet not overly restrictive strategy to analyze counterfactual distributions. Two sets of assumptions are analyzed here. First, the maximum (or minimum) monotone instrumental variables assumption, which imposes constraints on the variation of the potential outcome distribution conditional on the instrument. And second, the maximum (or minimum) monotone treatment response assumption, which restricts the shape of the outcome distribution over potential states. As a general result, point identification can only be achieved in special cases. However, if both types of assumptions are imposed jointly, then they may significantly improve over the no-assumptions bounds.

The bounding strategy poses two limits on the analysis of causal effects. First, it requires sufficiently large and rich data since the outcome distribution in each state is estimated separately. Furthermore, without additional assumptions it is generally not possible to extrapolate to outcomes not actually observed. Second, while in some cases it is possible to test the joint set of assumptions, and thus to evaluate their plausibility, the imposed structure may still be too weak to allow drawing informative conclusions regarding the stochastic order, even partially. It might therefore be desirable to combine the assumptions of this paper with additional restrictions, such as concavity or convexity, some sort of exclusion restriction, or additional information regarding the outcome mechanism.

A crucial issue in the construction of bounds is credible knowledge of the extremum points, which may not always exist, and it would be helpful to develop a data driven criterion to identify global minima or maxima. Further research is also needed regarding inference. First, the potential finite sample bias in the analogue estimates needs to be appropriately corrected (see Kreider and Pepper 2007, and Manski and Pepper 2009, for related results). Second, the uncertainty in the sign of the inequalities and the location of the extremum points in the MMIV/MMTS and the MMTR assumptions must be accounted for in order to derive confidence intervals with a pre-defined coverage probability.

References

Andrews, D.W.K., S. Berry, and P. Jia (2004), Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location, *unpublished manuscript*.

Andrews, D.W.K., and S. Han (2009), Invalidity of the Bootstrap and the m Out of n Bootstrap for Confidence Interval Endpoints Defined by Moment Inequalities, *The Econometrics Journal*, 12, S172-S199.

Andrews, D.W.K., and P. Guggenberger (2009), Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities, *Econometric Theory*, 25, 669-709.

Balke, A., and J. Pearl (1997), Bounds on Treatment Effects from Studies with Imperfect Compliance, *Journal of the American Statistical Association*, 92, 1171-1176.

Bhattacharya, J., A.M. Shaikh, and E. Vytlacil (2008), Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization, *American Economic Review*, 98, 351-356.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007), Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds, *Econometrica*, 75, 323-363.

Boes, S. (2009), Partial Identification of Discrete Counterfactual Distributions With Sequential Update of Information, *SOI WP 0918*.

Boes, S. (2010), Nonparametric Analysis of Treatment Effects in Ordered Response Models, *Empirical Economics*, forthcoming.

Broz, J.L, and M. Plouffe (2010), The Effectiveness of Monetary Policy Anchors: Firm-Level Evidence, *International Organization*, forthcoming.

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

Chesher, A. (2007), Endogeneity and Discrete Outcomes, *CEMMAP Working Paper CWP05/07.*

Chesher, A. (2008), Instrumental Variable Models for Discrete Outcomes, *CEMMAP Working Paper CWP30/08.*

Chiburis, R.C. (2009), Semiparametric Bounds on Treatment Effects, *unpublished manuscript.*

Cunha, F., J.J. Heckman, and S. Navarro (2008), The Identification And Economic Content Of Ordered Choice Models With Stochastic Thresholds, *International Economic Review*, 48, 1273-1309.

Currie, J., and E. Moretti (2003), Mothers Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings, *Quarterly Journal of Economics*, 118, 1495-1532.

Cutler, D., and Lleras-Muney, A. (2009), Understanding Differences in Health Behavior by Education, *Journal of Health Economics*, forthcoming.

de Haan, M. (2008), The Effect of Parents' Schooling on Child's Schooling: A Nonparametric Bounds Analysis, *unpublished manuscript.*

de Walque, D. (2007), Does Education Affect Smoking Behaviors? Evidence using the Vietnam Draft as an Instrument for College Education, *Journal of Health Economics*, 26, 877-895.

Ermisch, J.F., and M. Francesconi (2001), Family Structure and Children's Achievements, *Journal of Population Economics*, 14, 249-270.

Farrell, P., and V.R. Fuchs (1982), Schooling and Health: The Cigarette Connection, *Journal of Health Economics*, 1, 217-230.

Gerfin, M., and M. Schellhorn (2006), Nonparametric Bounds on the Effect of Deductibles in Health Care Insurance on Doctor VisitsSwiss Evidence, *Health Economics*, 15, 1011-1020.

Gilman, S.E., L.T. Martin, D.B. Abrams, I. Kawachi, L. Kubzansky, E.B. Loucks, R. Rende, R. Rudd, and S.L. Buka (2008), Educational Attainment and Cigarette Smoking: A Causal Association? *International Journal of Epidemiology*, 37, 615-624.

González, L. (2005), Nonparametric Bounds on the Returns to Language Skills, *Journal of Applied Econometrics*, 20, 771-795.

Grimard, F., and D. Parent (2007), Education and Smoking: Were Vietnam Draft Avoiders Also More Likely to Avoid Smoking? *Journal of Health Economics*, 26, 896-926.

Grossman, M. (1972), On the Concept of Health Capital and the Demand for Health, *Journal of Political Economy*, 80, 223-255.

Guggenberger P., J. Hahn, and K. Kim (2008), Specification Testing Under Moment Inequalities, *Economics Letters*, 99, 375-378.

Heckman, J.J., and E.J. Vytlacil (2001), Local Instrumental Variables, in: C. Hsiao, K. Morimune, and J. Powell (eds.) *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, Cambridge: Cambridge University Press, 1-46.

Imbens, G.W., and C.F. Manski (2004), Confidence Intervals for Partially Identified Parameters, *Econometrica*, 72, 1845-1857.

Kenkel, D.S. (1991), Health Behavior, Health Knowledge, and Schooling, *Journal of Political Economy*, 99, 287-305.

Kenkel, D.S., D. Lillard, and A. Mathios (2006), The Roles of High School Completion and GED Receipt in Smoking and Obesity, *Journal of Labor Economics*, 24, 635-660.

Kreider, B., and S. Hill (2009), Partially Identifying Treatment Effects with an Application to Covering the Uninsured, *Journal of Human Resources*, 44, 409-449.

Kreider, B., and J.V. Pepper (2007), Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors, *Journal of the American Statistical Association*, 102, 432-441.

Kreider, B., and J.V. Pepper (2008), Inferring Disability in Corrupt Data, *Journal of Applied Econometrics*, 23, 329-349.

Machado, C., A.M. Shaikh, and E. Vytlacil (2009), Instrumental Variables and the Sign of the Average Treatment Effect, *unpublished manuscript*.

Manski, C.F. (1990), Nonparametric Bounds on Treatment Effects, *American Economic Review Papers and Proceedings*, 80, 319-323.

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

Manski, C.F. (1997), Monotone Treatment Response, *Econometrica*, 65, 1311-1334.

Manski, C.F. (2003), *Partial Identification of Probability Distributions*, Springer.

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

Manski, C.F., and J.V. Pepper (2009), More on Monotone Instrumental Variables, *The Econometrics Journal*, 12, S200-S216.

Manski, C.F. and E. Tamer (2002), Inference on Regressions with Interval Data on a Regressor or Outcome, *Econometrica*, 70, 519-546.

Mullahy, J. (1985), *Cigarette Smoking: Habits, Health Concern, and Heterogeneous Unobservables in a Microeconometric Analysis of Consumer Demand. PhD Dissertation*, University of Virginia.

Okumura, T., and E. Usui (2007), Concave-Monotone Treatment Response and Monotone Treatment Selection: With Returns to Schooling Application, *unpublished manuscript.*

Pakes, A. (2009), Alternative Models for Moment Inequalities, *manuscript.*

Pakes A., J. Porter, K. Ho, and J. Ishii (2006), Moment Inequalities and Their Application, *manuscript.*

Romano, J.P., and A.M. Shaikh (2009), Inference for the Identified Set in Partially Identified Econometric Models, *Econometrica*, forthcoming.

Rosen, A. (2008), Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities, *Journal of Econometrics*, 146, 107-117.

Rosenzweig, M.R. (1995), Why are there Returns to Schooling? *American Economic Review*, 85, 153-58.

Shaikh, A.M. and E.J. Vytlacil (2005), Threshold Crossing Models and Bounds on Treatment Effects, *NBER Technical Working Paper No. 307.*

Stoye, J. (2009), More on Confidence Regions for Partially Identified Parameters, *Econometrica*, forthcoming.

Tamer, E. (2003), Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, *Review of Economic Studies*, 70, 147-165.

Tamer, E. (2009), Partial Identification in Econometrics, *manuscript.*

Tenn, S., D.A. Herman, and B. Wendling (2008), The Role of Education in the Production of Health: An Empirical Analysis of Smoking Behavior, FTC Bureau of Economics Working Paper No. 292.

Wildman, J., A.M. Jones (2008), Disentangling the Relationship Between Health and Income, *Journal of Health Economics*, 27, 308-324.

Tables and Figures

Table 1: Smoking distribution conditional on schooling

	Schooling categories (years)				
	≤ 8	9-11	12	13-15	≥ 16
P(S)	0.090	0.136	0.343	0.172	0.259
	[178]	[270]	[681]	[340]	[512]
$P(cigs = 0\|S)$	0.523	0.485	0.584	0.627	0.760
$P(cigs \leq 1/2\ pack\|S)$	0.618	0.548	0.648	0.688	0.799
$P(cigs \leq 1\ pack\|S)$	0.781	0.752	0.824	0.835	0.883
$P(cigs \leq 1\,1/2\ packs\|S)$	0.860	0.856	0.888	0.915	0.920

Source: Smoking Supplement of the 1979 US National Health Interview Survey, own calculations. *Notes:* The estimates are based on a random sample of 1,981 observations on employed white men aged 35-60. The numbers of observations used to estimate each cumulative distribution are reported in square brackets. *cigs* denotes the number of cigarette packs smoked per day, S denotes the years of schooling in categories as shown.

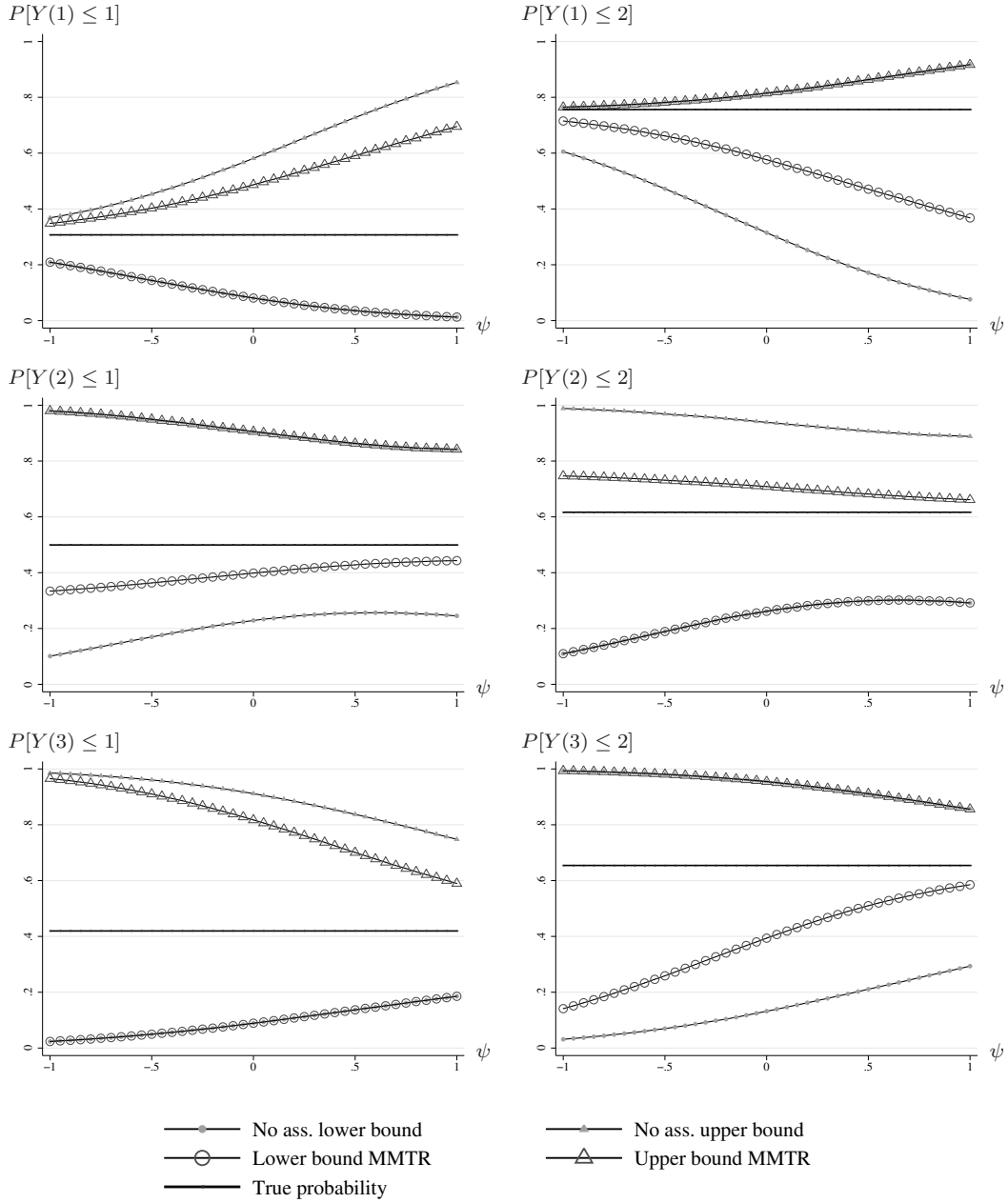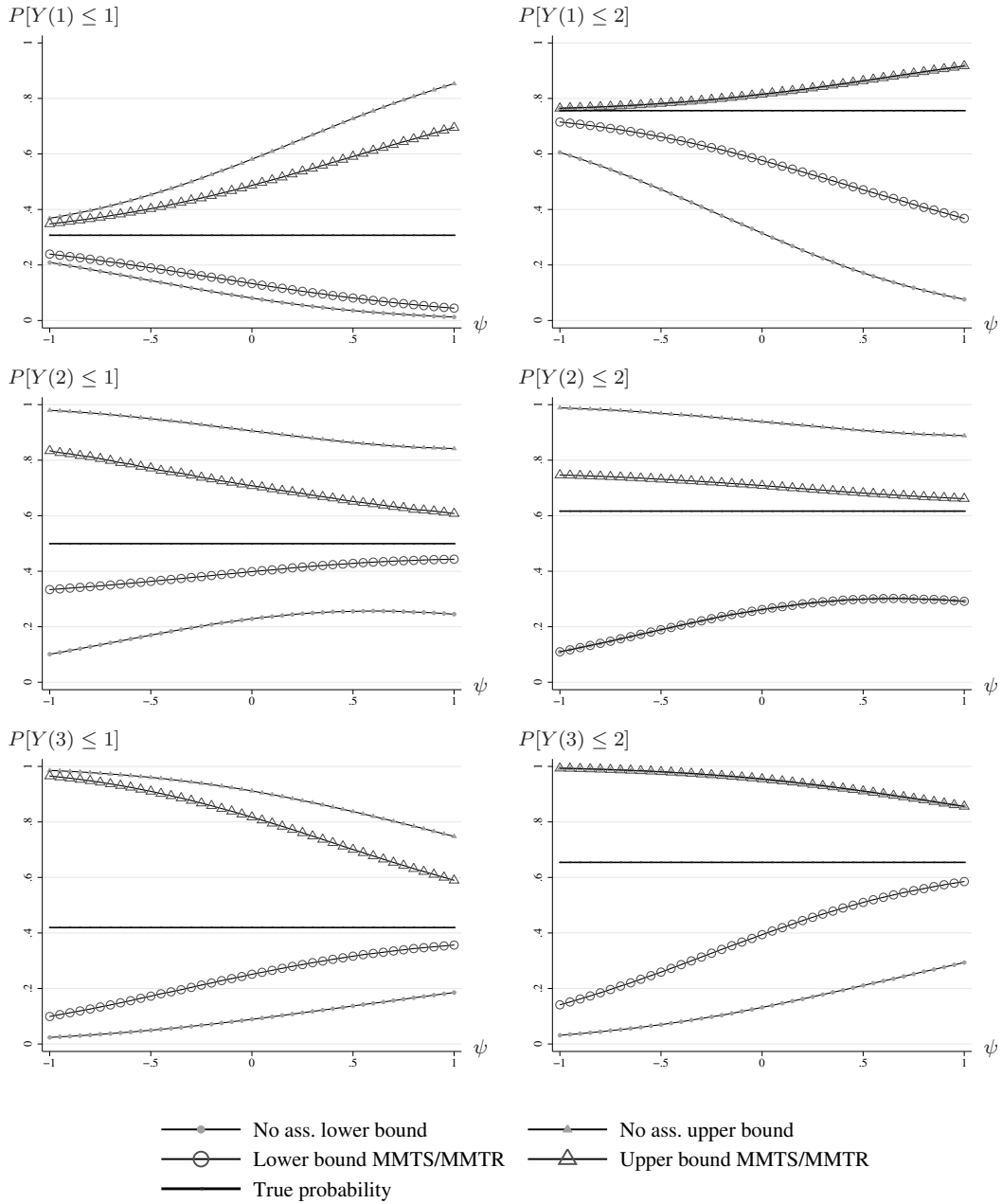Figure 1: Bounds under MMTS only – Scenario 1
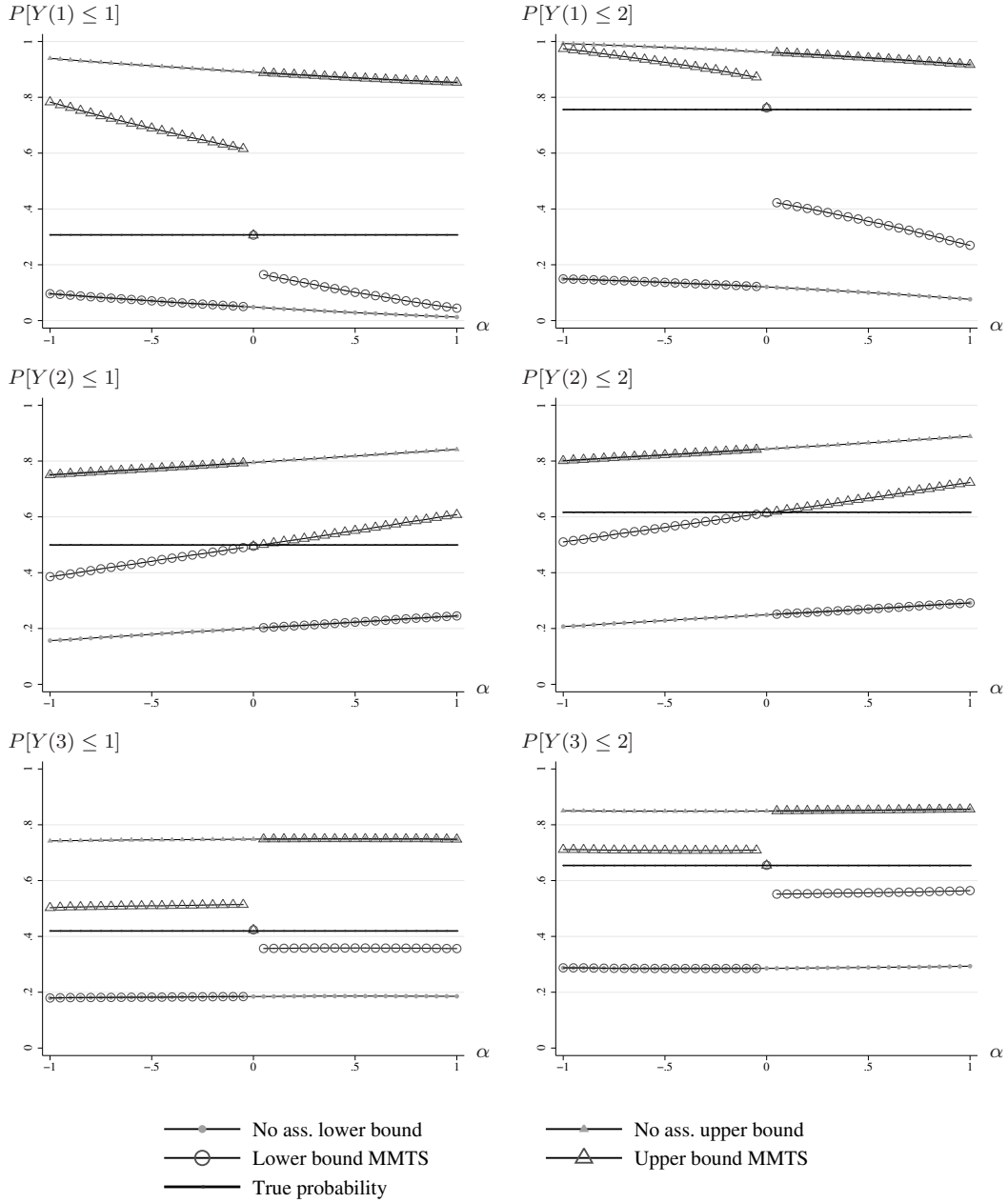
Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Corollary 1 under the MMTS assumption only.
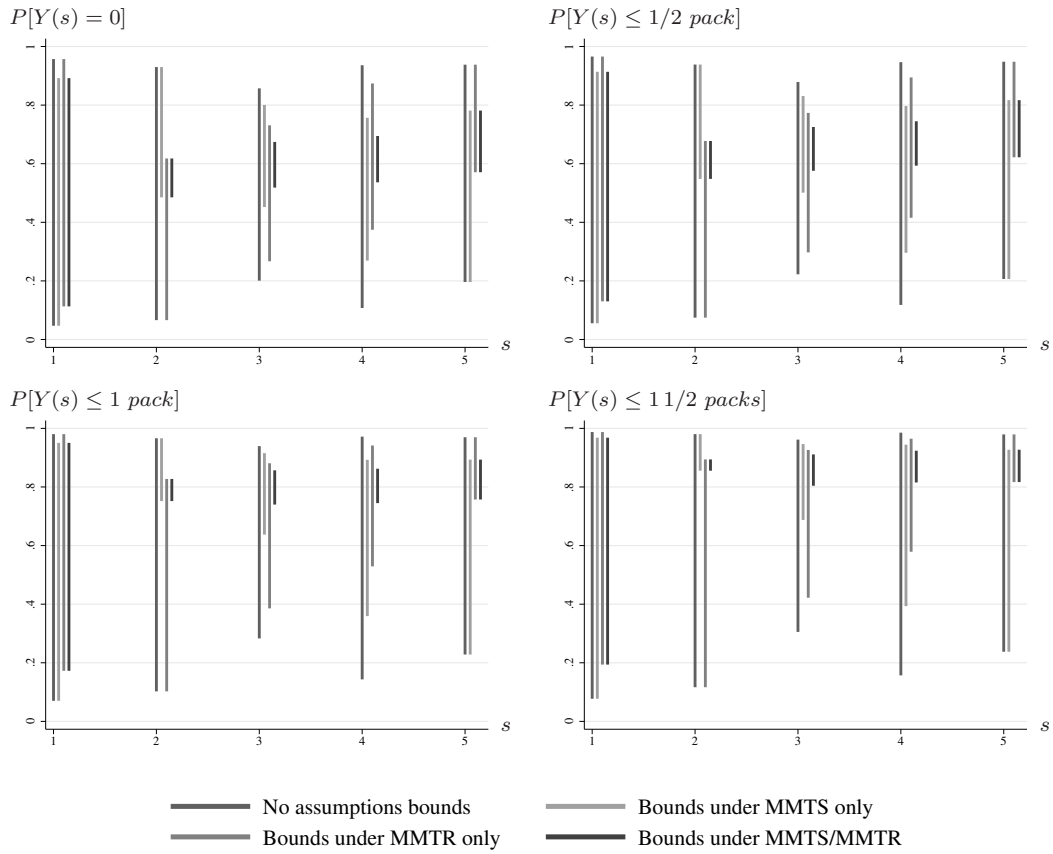
Figure 2: Bounds under MMTR only – Scenario 1
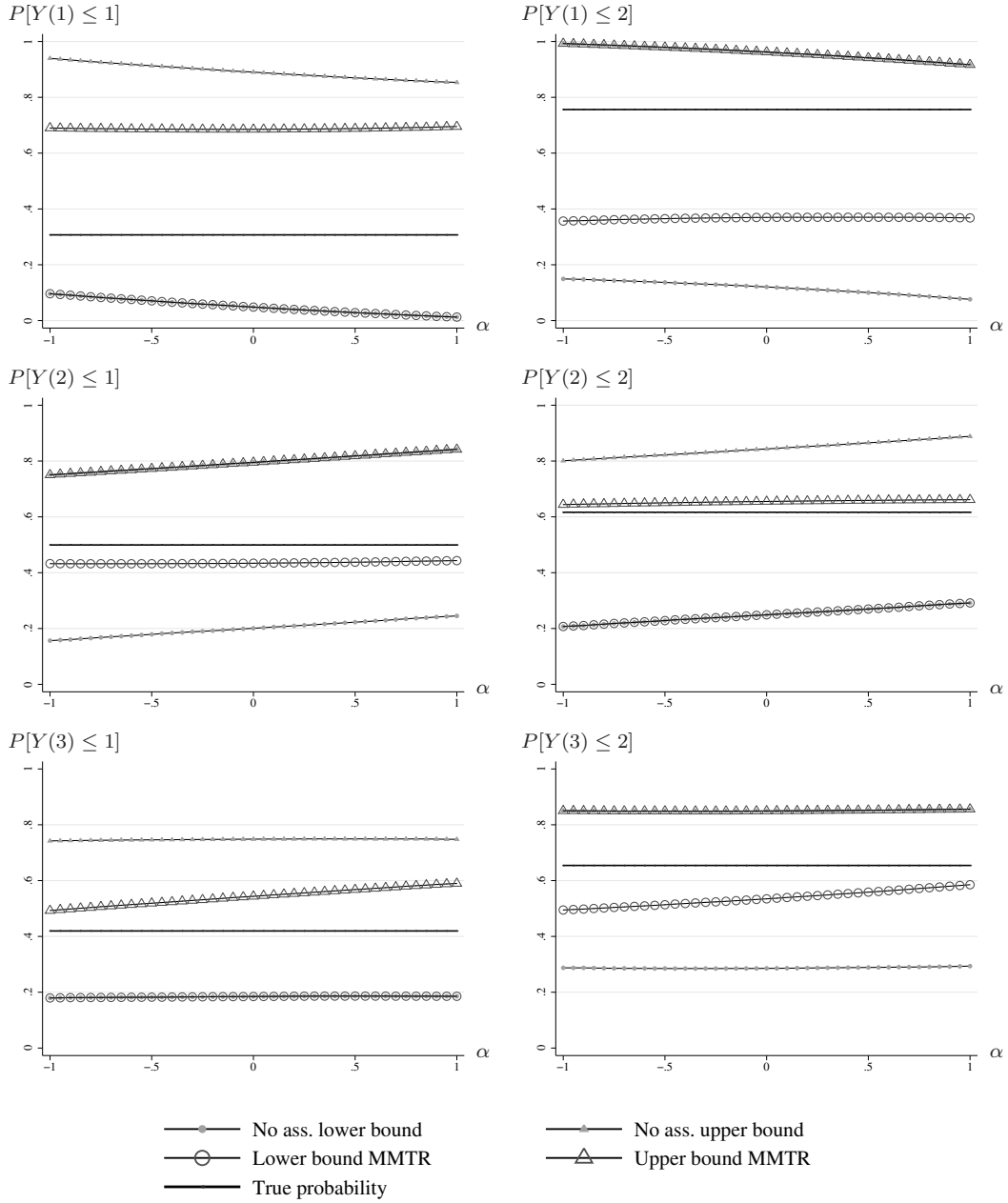
Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 3 under the MMTR assumption only.

Figure 3: Bounds under MMTS/MMTR assumptions – Scenario 1
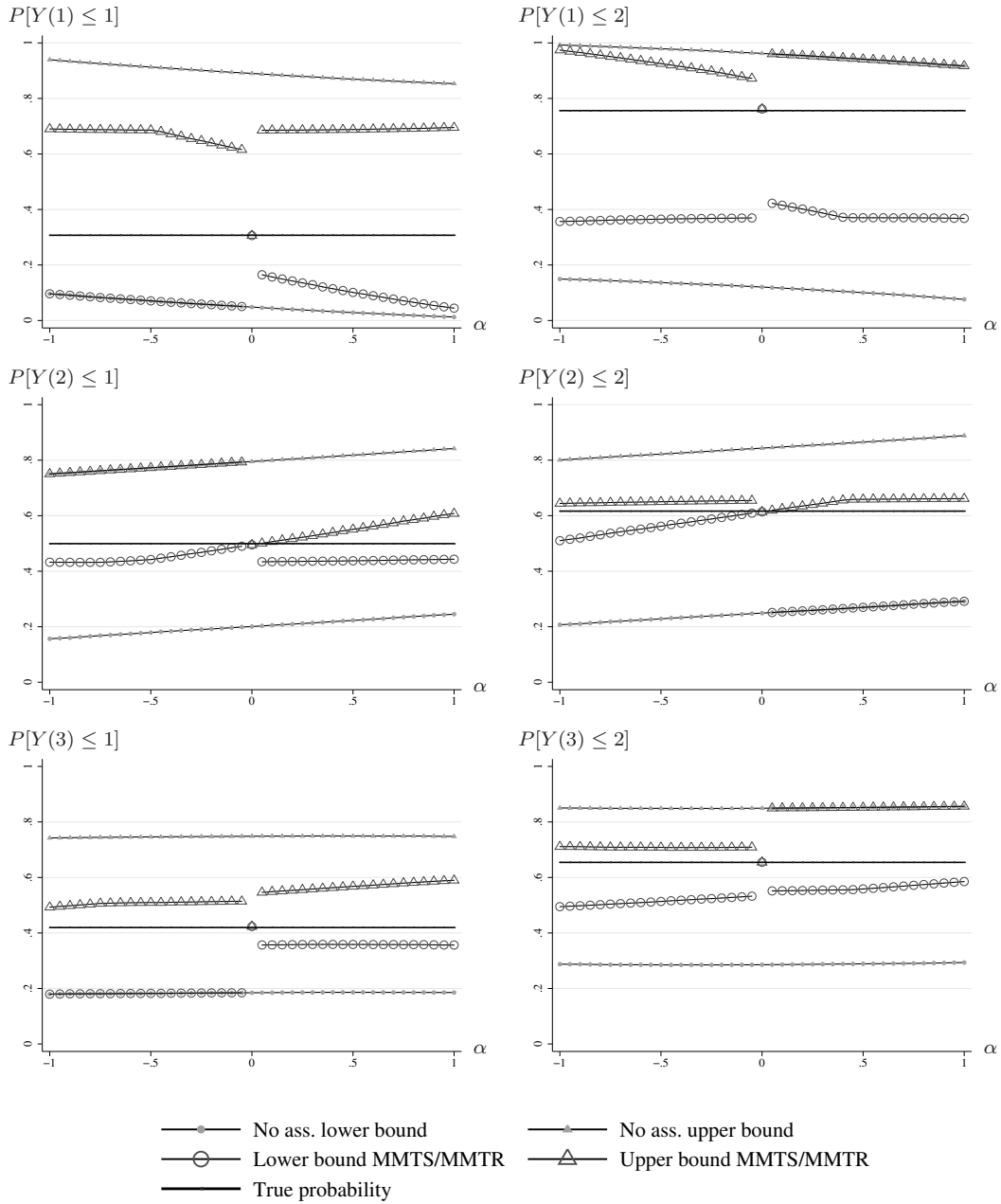
Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds under the MMTS/MMTR assumptions.

31

Figure 4: Bounds under MMTS only – Scenario 2
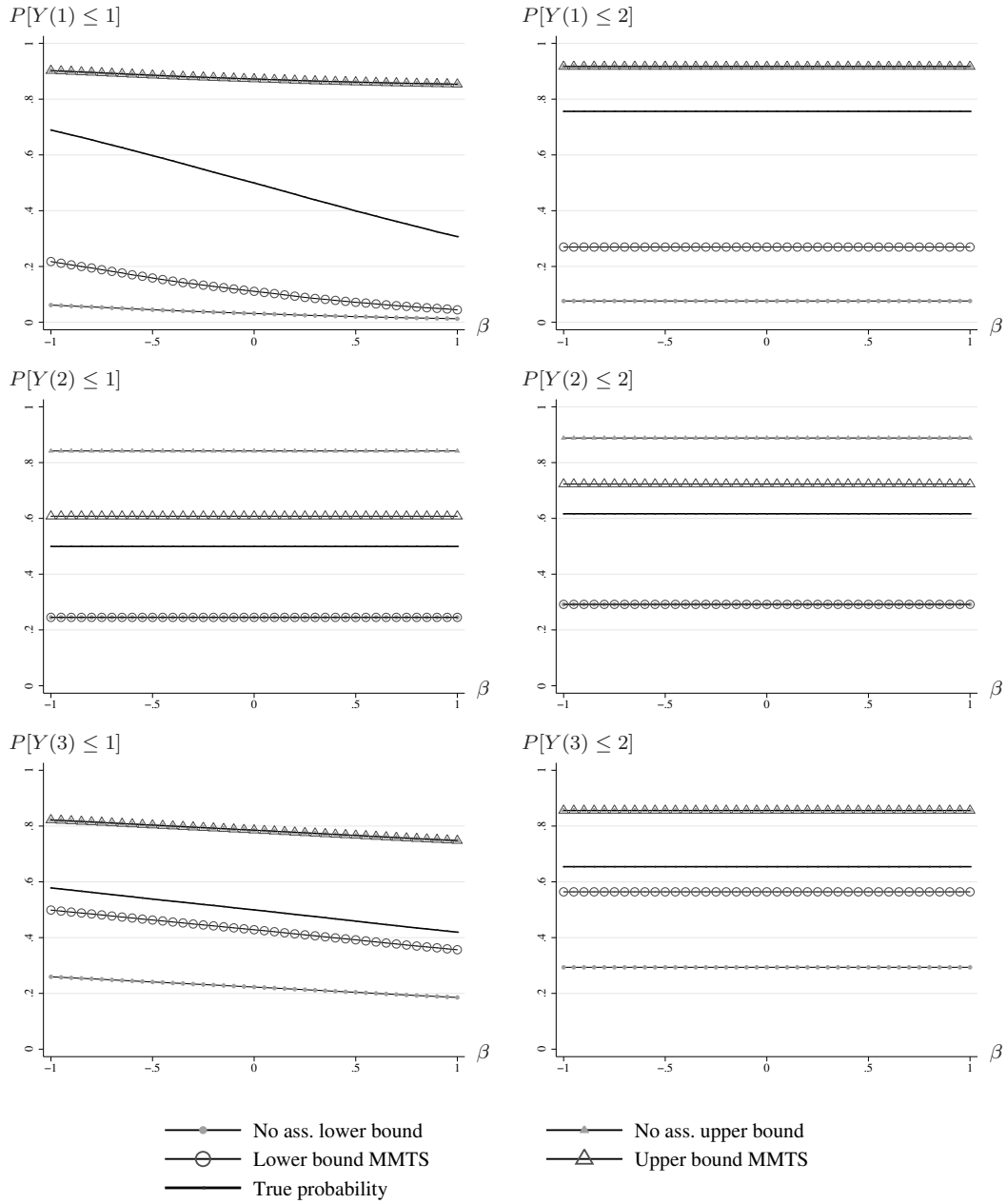
Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Corollary 1 under the MMTS assumption only.

Figure 5: Bounds under MMTR only – Scenario 2
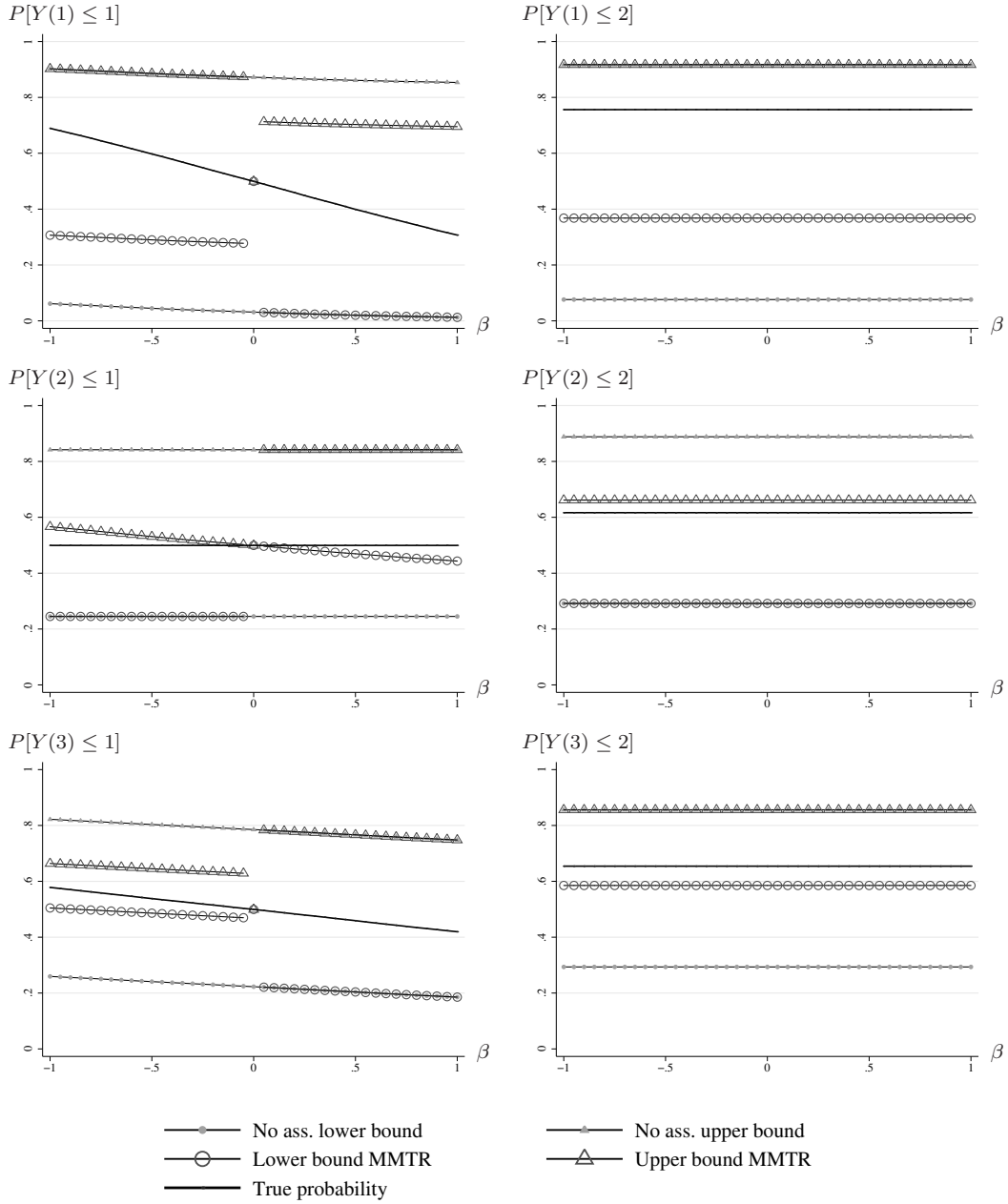
$P[Y(1) \leq 1]$

$P[Y(1) \leq 2]$

$P[Y(2) \leq 1]$

$P[Y(2) \leq 2]$

$P[Y(3) \leq 1]$

$P[Y(3) \leq 2]$

No ass. lower bound   No ass. upper bound
Lower bound MMTR   Upper bound MMTR
True probability

Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 3 under the MMTR assumption only.

Figure 6: Bounds under MMTS/MMTR assumptions – Scenario 2
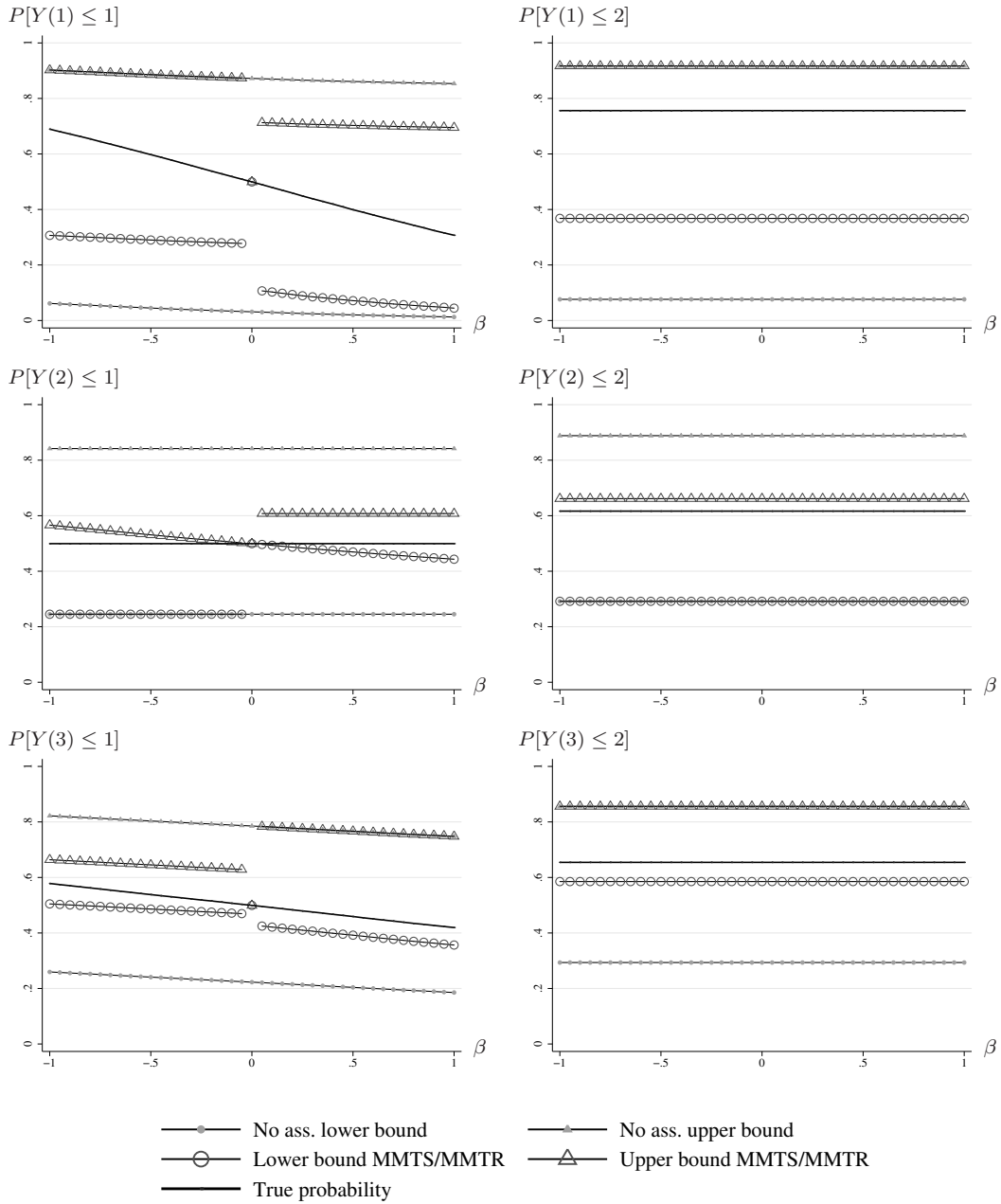
Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds under the MMTS/MMTR assumptions.

Figure 7: Bounds under MMTS only – Scenario 3
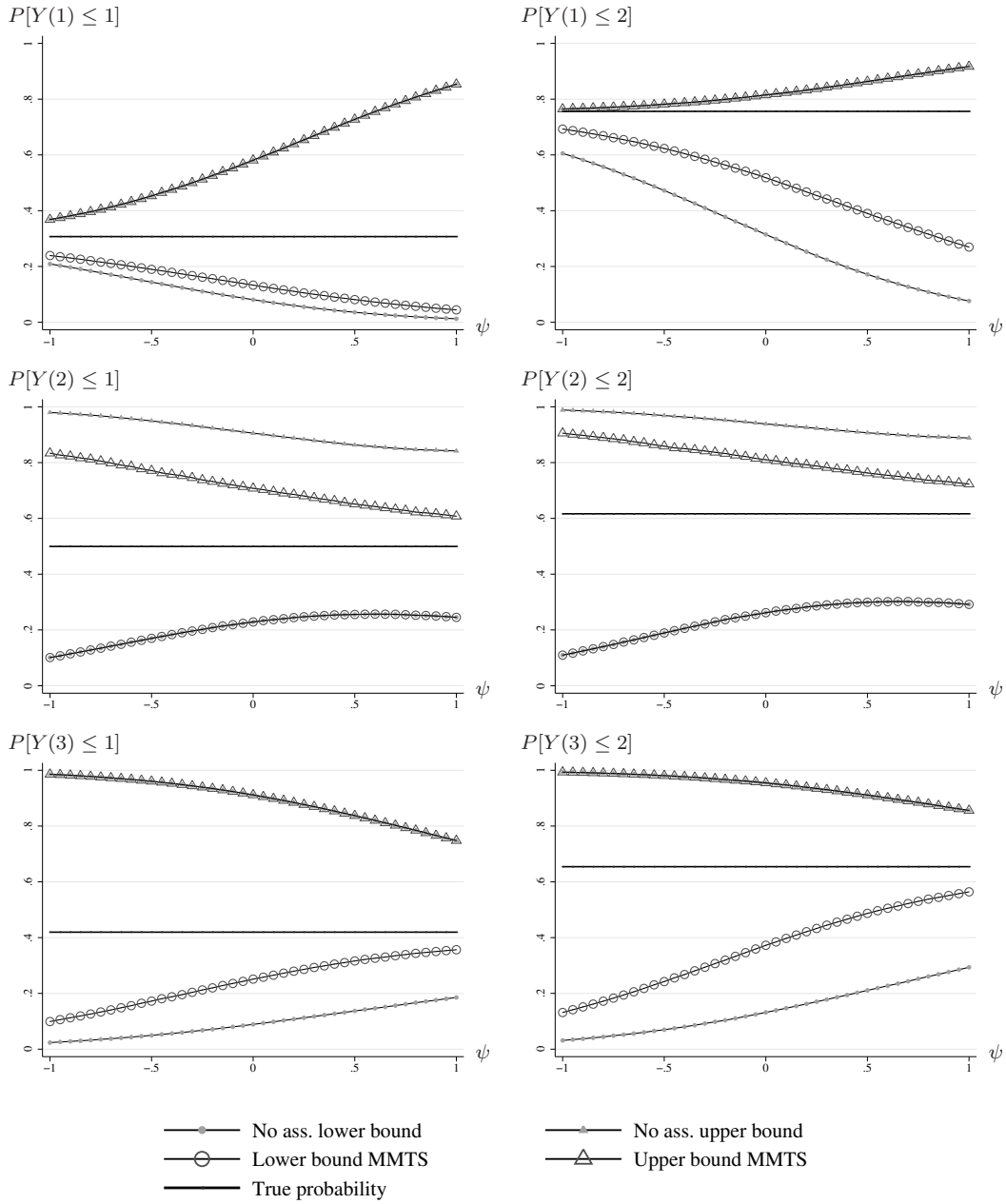
Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Corollary 1 under the MMTS assumption only.

Figure 8: Bounds under MMTR only – Scenario 3

Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 3 under the MMTR assumption only.

Figure 9: Bounds under MMTS/MMTR assumptions – Scenario 3

Notes: Thick black lines denote the true potential outcome distribution $P[Y(s) \leq y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds under the MMTS/MMTR assumptions.

Figure 10: Bounds on the potential smoking distribution by schooling level

$P[Y(s) = 0]$

$P[Y(s) \leq 1/2 \ pack]$

$P[Y(s) \leq 1 \ pack]$

$P[Y(s) \leq 1\ 1/2 \ packs]$

No assumptions bounds   Bounds under MMTS only
Bounds under MMTR only   Bounds under MMTS/MMTR

Source: Smoking Supplement of the 1979 US National Health Interview Survey, own calculations. *Notes:* The estimates are based on a random sample of 1,981 observations on employed white men aged 35-60. Left bars are obtained using the empirical evidence alone. The MMTS assumption asserts that $2 = \arg\min_{u \in \mathcal{S}} P[Y(s) \leq y | S = u]$ (middle left bars), the MMTR assumption asserts that $2 = \arg\min_{s \in \mathcal{S}} P[Y(s) \leq y | S = u]$ (middle right bars). The right bars indicate the bounds obtained under the joint set of assumptions MMTS and MMTR.

Working Papers of the Socioeconomic Institute at the University of Zurich

0920	Bounds on Counterfactual Distributions Under Semi-Monotonicity Constraints, Stefan Boes, December 2009, 38 p.
0919	Rotten Kids with Bad Intentions, Nick Netzer, Armin Schmutzler, December 2009, 38 p.
0918	Partial Identification of Discrete Counterfactual Distributions with Sequential Update of Information, Stefan Boes, December 2009, 37 p.
0917	How much do journal titles tell us about the academic interest and relevance of economic reserach? An empirical analysis, Felix Schläpfer, December 2009, 14 p.
0916	Fine Tuning of Health Insurance Regulation: Unhealthy Consequences for an Individual Insurer, Johannes Schoder, Michèle Sennhauser, Peter Zweifel, August 2009, 18 p.
0915	Capping Risk Adjustment?, Patrick Eugster, Michèle Sennhauser, Peter Zweifel, September 2009, 27 p.
0914	A Pharmaceutical Innovation: Is it Worth the Money? Whose Money?, Michèle Sennhauser, Peter Zweifel, September 2009, 22 p.
0913	Copula-based bivariate binary response models, Rainer Winkelmann, August 2009, 26 p.
0912	Simulating WTP Values from Random-Coefficient Models, Maurus Rischatsch, July 2009, 6 p.
0911	Physician dispensing and the choice between generic and brand-name drugs – Do margins affect choice?, Maurus Rischatsch, Maria Trottmann, July 2009, 15 p.
0910	GPs' preferences: What price fee-for-service?, Peter Zweifel, Maurus Rischatsch, Angelika Brändle, July 2009, 21 p.
0909	Social Mobility and Preferences for Income Redistribution: Evidence from a Discrete Choice Experiment, Ilja Neustadt, Peter Zweifel, July 2009, 31 p.
0908	Robust estimation of zero-inflated count models, Kevin E. Staub, Rainer Winkelmann, June 2009, 22 p.
0907	Competitive Screening in Insurance Markets with Endogenous Wealth Heterogeneity, Nick Netzer, Florian Scheuer, April 2009, 28 p.
0906	New Flight Regimes and Exposure to Aircraft Noise: Identifying Housing Price Effects Using a Ratio-of-Ratios Approach, Stefan Boes, Stephan Nüesch, April 2009, 40 p.
0905	Patents versus Subsidies – A Laboratory Experiment, Donja Darai, Jens Großer, Nadja Trhal, March 2009, 59 p.
0904	Simple tests for exogeneity of a binary explanatory variable in count data regression models, Kevin E. Staub, February 2009, 30 p.
0903	Spurious correlation in estimation of the health production function: A note, Sule Akkoyunlu, Frank R. Lichtenberg, Boriss Siliverstovs, Peter Zweifel, February 2009, 13 p.
0902	Making Sense of Non-Binding Retail-Price Recommendations, Stefan Bühler, Dennis L. Gärtner, February 2009, 30 p.
0901	Flat-of-the-Curve Medicine – A New Perspective on the Production of Health, Johnnes Schoder, Peter Zweifel, January 2009, 35 p.
0816	Relative status and satisfaction, Stefan Boes, Kevin E. Staub, Rainer Winkelmann, December 2008, 11 p.
0815	Delay and Deservingness after Winning the Lottery, Andrew J. Oswald, Rainer Winkelmann, December 2008, 29 p.