University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 1001

Convex Treatment Response and Treatment Selection

Stefan Boes

January 2010

Socioeconomic Institute
University of Zurich

Working Paper No. 1001

Convex Treatment Response and Treatment Selection

January 2010

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

Publisher
Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-44-634 21 37
Fax: +41-44-634 49 82
URL: www.soi.uzh.ch
E-mail: soilib@soi.uzh.ch

Convex Treatment Response and Treatment Selection*

Stefan Boes†

January 16, 2010

Abstract

This paper analyzes the identifying power of weak convexity assumptions in treatment effect models with endogenous selection. The counterfactual distributions are constrained either in terms of the response function, or conditional on the realized treatment, and sharp bounds on the potential outcome distributions are derived. The methods are applied to bound the effect of education on smoking.

JEL Classification: C14, C30, I12

Keywords: Nonparametric bounds, causality, endogeneity, instrumental variables.

---

*Acknowledgments: Financial support was generously provided through Swiss NSF grant #PBZHP1-124191. I thank Karim Chalak, Gary Chamberlain, Rustam Ibragimov, Guido Imbens, Gary King, Chris Winship, and seminar participants at Harvard for valuable comments. John Mullahy kindly provided the data.

†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

The credibility of empirical work crucially depends on the strength of the underlying assumptions (Manski 2003). Stronger assumptions allow to draw stronger conclusions, but there is generally less consensus about their validity. This paper avoids assumptions like conditional independence, exogenous treatment selection, or exclusion restrictions that are commonly employed in treatment effect models. The paper builds on the previous literature on monotone treatment response (e.g., Manski 1997; Blundell *et al.* 2007; Okumura and Usui 2009) and monotone instrumental variables (e.g., Manski and Pepper 2000, 2009; Boes 2009) by introducing weak convexity (or concavity) in the counterfactual distributions.

To formalize the discussion, let each individual have a response function $Y(s) \in \mathcal{Y}$ that determines the outcome in state $s \in \mathcal{S}$. Assume that outcomes are measured (at least) on the ordinal scale, and states are measured on the interval scale. Let $S \in \mathcal{S}$ denote the realized state, let $Y = Y(S)$ denote the realized outcome, and let $X \in \mathcal{X}$ denote the vector of covariates. The observed data are the triple (Y, S, X) with distribution $P(Y, S, X)$.

The goal is to learn the potential outcome distributions $P[Y(s) \leq y]$ or $P[Y(s) \leq y|X]$. Two fundamental problems impede the point identification of $P[Y(s) \leq y]$ (covariates are kept implicit in what follows to simplify notation). First, the outcome of each individual is only observed in the realized state, outcomes that would be realized under alternative states are logically unobserved. Second, the distribution of outcomes observed does not necessarily resemble the distribution of potential outcomes (e.g., due to self-selection).

The paper imposes weak restrictions on the curvature of the potential outcome distributions, either conditional on the treatment indicator, or in terms of the response function. More specifically, $P[Y(s) \leq y|S = t]$ is assumed to be convex/concave in t and/or in s, which will be referred to as convex/concave treatment selection and convex/concave treatment response, respectively. While each assumption is non-refutable by the empirical evidence alone, the joint assumptions are refutable. Moreover, the bounds under convexity/concavity may yield a substantial improvement over the no-assumptions bounds.

2 Assumptions and Bounds

2.1 Convex/Concave Treatment Selection

The first type of assumption restricts the curvature of the potential outcome distribution as a function of realized states, i.e., conditional on different values of S. Any linear combination of $P[Y(s) \leq y|S]$ evaluated at $S = t_1$ and evaluated at $S = t_2$, with $t_1 \leq t_2$ and combination coefficient α, is assumed to weakly dominate the distribution evaluated at $S = \alpha t_1 + (1-\alpha)t_2$. This assumption is referred to as convex treatment selection (CXTS), formally

Assumption (CXTS)***.*** *For each* $y \in \mathcal{Y}$*,* $(s, t_1, t_2) \in \mathcal{S}^3$*, with* $t_1 \leq t_2$*, and for all* $\alpha \in [0, 1]$*, let* $P[Y(s) \leq y|S = \alpha t_1 + (1-\alpha)t_2] \;\leq\; \alpha P[Y(s) \leq y|S = t_1] + (1-\alpha)P[Y(s) \leq y|S = t_2]$.

Assumption CXTS has two implications regarding the potential outcome distribution. First, for any $t \in \mathcal{S}$ with $t_1 \leq t \leq t_2$ and evaluation points $(t_1, t_2) \in \mathcal{S}^2$, and $\forall y \in \mathcal{Y}$, there exists α^* (as a function of t_1 and t_2) with $t = \alpha^* t_1 + (1-\alpha^*)t_2$ such that

$$P[Y(s) \leq y|S = t] \;\leq\; \alpha^* P[Y(s) \leq y|S = t_1] + (1-\alpha^*)P[Y(s) \leq y|S = t_2]. \tag{1}$$

Thus, an upper bound of the right-hand side is also an upper bound of the left-hand side. Since the weak inequality in (1) holds for any $t_1 \leq t$ and any $t_2 \geq t$, the smallest of the upper bounds over all (t_1, t_2) can be used as sharp upper bound for $P[Y(s) \leq y|S = t]$.

Second, any convex function exhibits one of three types of (weak) monotonicity: (i) monotonically decreasing in S, (ii) monotonically increasing in S, or (iii) one switch from monotonically decreasing to monotonically increasing with a minimum at $t_{\min} = \arg\min_{t \in \mathcal{S}} P[Y(s) \leq y|S = t]$. The former two cases can be subsumed in the third if $t_{\min} = t_{(\min)}$ and $t_{\min} = t_{(\max)}$, respectively, where $t_{(\min)}$ denotes the smallest point and $t_{(\max)}$ the largest point in $\mathcal{S}$ (the existence of these points is ensured by the assumption that S is measured on the interval scale). Define $F(y|s) = P(Y \leq y|S = s)$, $\mathcal{S}^l_{t,\min} = \{t \in \mathcal{S} : t \leq t_{\min}\}$, and $\mathcal{S}^u_{t,\min} = \{t \in \mathcal{S} : t \geq t_{\min}\}$. The bounds on the counterfactual distributions can be summarized as follows:

Proposition 1. *Let assumption CXTS hold. If* $t = s$*, then* $P[Y(s) \leq y|S = t] = F(y|s)$.

If $s < t_{\min}$*, then*

$$\begin{array}{llll} t \in \mathcal{S}^l_{t,\min} & t < s & \Rightarrow & F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) + \dfrac{s - t}{s - t_{(\min)}}[1 - F(y|s)] \\ & t > s & \Rightarrow & 0 \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) \\ t \notin \mathcal{S}^l_{t,\min} & & \Rightarrow & 0 \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 - \dfrac{t_{(\max)} - t}{t_{(\max)} - t_{\min}}[1 - F(y|s)] \end{array}$$

If $s = t_{\min}$*, then*

$$\begin{array}{lll} t < s & \Rightarrow & F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) + \dfrac{t_{\min} - t}{t_{\min} - t_{(\min)}}[1 - F(y|s)] \\ t > s & \Rightarrow & F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 - \dfrac{t_{(\max)} - t}{t_{(\max)} - t_{\min}}[1 - F(y|s)] \end{array}$$

If $s > t_{\min}$*, then*

$$\begin{array}{llll} t \in \mathcal{S}^u_{t,\min} & t < s & \Rightarrow & 0 \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) \\ & t > s & \Rightarrow & F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 - \dfrac{t_{(\max)} - t}{t_{(\max)} - s}[1 - F(y|s)] \\ t \notin \mathcal{S}^u_{t,\min} & & \Rightarrow & 0 \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) + \dfrac{t_{\min} - t}{t_{\min} - t_{(\min)}}[1 - F(y|s)] \end{array}$$

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

Proof. See the appendix.

The CXTS assumption yields point identification if $t < s$, t and s both lie in the monotonically decreasing part of $P[Y(s) \leq y|S = t]$ over realized states, and $t_{(\min)}$ approaches $-\infty$ in the limit, or if $t > s$, t and s both lie in the monotonically increasing part of $P[Y(s) \leq y|S = t]$ over realized states, and $t_{(\max)}$ approaches ∞ in the limit. Since the target is a distribution function, with natural lower and upper limits of zero and one, point identification will only be achieved for an idealized shape of the counterfactual distributions conditional on S.

Proposition 1 can be applied to derive bounds on the potential outcome distributions:

Corollary 1. *Let assumption CXTS hold. Then,*

$$
\begin{aligned}
s < t_{\min} \;\Rightarrow\; & F(y|s)P(S \le s) \;\le\; P[Y(s) \le y] \;\le\; F(y|s)P(S \le t_{\min}) \\
& + \sum_{t<s} \left(\frac{s-t}{s-t_{(\min)}}[1-F(y|s)] \right) P(S=t) \\
& + \sum_{t>t_{\min}} \left(1 - \frac{t_{(\max)}-t}{t_{(\max)}-t_{\min}}[1-F(y|s)] \right) P(S=t) \\
s = t_{\min} \;\Rightarrow\; & F(y|s) \;\le\; P[Y(s) \le y] \;\le\; F(y|s)P(S=s) \\
& + \sum_{t<s} \left(F(y|s) + \frac{t_{\min}-t}{t_{\min}-t_{(\min)}}[1-F(y|s)] \right) P(S=t) \\
& + \sum_{t>s} \left(1 - \frac{t_{(\max)}-t}{t_{(\max)}-t_{\min}}[1-F(y|s)] \right) P(S=t) \\
s > t_{\min} \;\Rightarrow\; & F(y|s)P(S \ge s) \;\le\; P[Y(s) \le y] \;\le\; F(y|s)P(S \le s) \\
& + \sum_{t<t_{\min}} \left(\frac{t_{\min}-t}{t_{\min}-t_{(\min)}}[1-F(y|s)] \right) P(S=t) \\
& + \sum_{t>s} \left(1 - \frac{t_{(\max)}-t}{t_{(\max)}-s}[1-F(y|s)] \right) P(S=t)
\end{aligned}
$$

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

Proof. See the appendix.

Depending upon the data and the application as well as the implications of an underlying theoretical model, assumption CXTS might also be replaced by a weak concavity assumption:

Assumption (CVTS)**.** *For each $y \in \mathcal{Y}$, $(s, t_1, t_2) \in \mathcal{S}^3$, with $t_1 \le t_2$, and for all $\alpha \in [0,1]$, let $P[Y(s) \le y|S = \alpha t_1 + (1-\alpha)t_2] \;\ge\; \alpha P[Y(s) \le y|S = t_1] + (1-\alpha)P[Y(s) \le y|S = t_2]$.*

Assumption CVTS (the concave treatment selection assumption) states that any linear combination of the potential outcome distribution evaluated at $S = t_1$ and $S = t_2$, with combination coefficient α, is weakly dominated by the potential outcome distribution evaluated at $S = \alpha t_1 + (1-\alpha)t_2$. Define $t_{\max} = \arg\max_{t\in\mathcal{S}} P[Y(s) \le y|S = t]$, $\mathcal{S}^l_{t,\max} = \{t \in \mathcal{S} : t \le t_{\max}\}$, and $\mathcal{S}^u_{t,\max} = \{t \in \mathcal{S} : t \ge t_{\max}\}$. Then using analogous arguments as under convexity yields the following bounds on the counterfactual distributions:

Proposition 2. *Let assumption CVTS hold. If $t = s$, then $P[Y(s) \leq y|S = t] = F(y|s)$. If $s < t_{\max}$, then*

$$\begin{array}{lll} t \in \mathcal{S}^l_{t,\max} & t < s & \Rightarrow \dfrac{t - t_{(\min)}}{s - t_{(\min)}} F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) \\ & t > s & \Rightarrow F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 \\ t \notin \mathcal{S}^l_{t,\max} & & \Rightarrow \dfrac{t_{(\max)} - t}{t_{(\max)} - t_{\max}} F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 \end{array}$$

If $s = t_{\max}$, then

$$\begin{array}{ll} t < s & \Rightarrow \dfrac{t - t_{(\min)}}{t_{\max} - t_{(\min)}} F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) \\ t > s & \Rightarrow \dfrac{t_{(\max)} - t}{t_{(\max)} - t_{\max}} F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) \end{array}$$

If $s > t_{\max}$, then

$$\begin{array}{lll} t \in \mathcal{S}^u_{t,\max} & t < s & \Rightarrow F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 \\ & t > s & \Rightarrow \dfrac{t_{(\max)} - t}{t_{(\max)} - s} F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ F(y|s) \\ t \notin \mathcal{S}^u_{t,\max} & & \Rightarrow \dfrac{t - t_{(\min)}}{t_{\max} - t_{(\min)}} F(y|s) \ \leq \ P[Y(s) \leq y|S = t] \ \leq \ 1 \end{array}$$

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

Proof. See the appendix.

While the convexity assumption yields more informative upper bounds than they are obtained under monotonicity alone, assumption CVTS can be used to construct lower bounds on the counterfactual distributions that are more informative than the no-assumptions bounds and the bounds under monotonicity.

The bounds on the potential outcome distribution can be derived as follows:

Corollary 2. *Let assumption CVTS hold. Then,*

$$s < t_{\max} \ \Rightarrow \ F(y|s)P(s \leq S \leq t_{\max}) + \sum_{t<s} \left(\frac{t - t_{(\min)}}{s - t_{(\min)}} F(y|s) \right) P(S = t) + \sum_{t>t_{\max}} \left(\frac{t_{(\max)} - t}{t_{(\max)} - t_{\max}} F(y|s) \right) P(S = t)$$

$$
\begin{aligned}
&\leq\; P[Y(s)\leq y] \;\leq\; F(y|s)P(S\leq s)+P(S>s)\\
s=t_{\max}\quad\Rightarrow\quad & F(y|s)P(S=s)+\sum_{t<s}\left(\frac{t-t_{(\min)}}{t_{\max}-t_{(\min)}}F(y|s)\right)P(S=t)\\
&\sum_{t>s}\left(\frac{t_{(\max)}-t}{t_{(\max)}-t_{\max}}F(y|s)\right)P(S=t)\\
&\leq\; P[Y(s)\leq y] \;\leq\; F(y|s)\\
s>t_{\max}\quad\Rightarrow\quad & F(y|s)P(t_{\max}\leq S\leq s)+\sum_{t<t_{\max}}\left(\frac{t-t_{(\min)}}{t_{\max}-t_{(\min)}}F(y|s)\right)P(S=t)\\
&+\sum_{t>s}\left(\frac{t_{(\max)}-t}{t_{(\max)}-s}F(y|s)\right)P(S=t)\\
&\leq\; P[Y(s)\leq y] \;\leq\; F(y|s)P(S\geq s)+P(S<s)
\end{aligned}
$$

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

Proof. See the appendix.

2.2 Convex/Concave Treatment Response

The second type of assumption restricts the curvature of the potential outcome distribution in terms of the response function, i.e., over hypothetical states s. Any linear combination of the distribution in state s_1 and state $s_2 \geq s_1$, with combination coefficient β and for any given realized $S = t$, is assumed to weakly dominate the distribution in state $\beta s_1 + (1-\beta)s_2$. This assumption is referred to as convex treatment response (CXTR), formally

Assumption (CXTR)***.*** *For each $y \in \mathcal{Y}$, $(s_1, s_2, t) \in \mathcal{S}^3$, with $s_1 \leq s_2$, and for all $\beta \in [0,1]$, let $P[Y(\beta s_1 + (1-\beta)s_2) \leq y|S=t] \;\leq\; \beta P[Y(s_1) \leq y|S=t] + (1-\beta)P[Y(s_2) \leq y|S=t]$.*

Assumption CXTR has two implications regarding the potential outcome distribution which follow in close analogy to the convex treatment selection assumption. First, for any $s \in \mathcal{S}$ with $s_1 \leq s \leq s_2$ and hypothetical states $(s_1, s_2) \in \mathcal{S}^2$, and $\forall y \in \mathcal{Y}$, there exists β^* (as a function of s_1 and s_2) with $s = \beta^* s_1 + (1-\beta^*)s_2$ such that

$$
P[Y(s)\leq y|S=t] \;\leq\; \beta^* P[Y(s_1)\leq y|S=t]+(1-\beta^*)P[Y(s)\leq y|S=t]. \tag{2}
$$

Thus, an upper bound of the right-hand side is also an upper bound of the left-hand side. Since the weak inequality in (2) holds for any $s_1 \leq s$ and any $s_2 \geq s$, the smallest of the upper bounds over all (s_1, s_2) can be used as sharp upper bound for $P[Y(s) \leq y|S = t]$.

Second, any convex function exhibits one of three types of (weak) monotonicity: (i) monotonically decreasing in s, (ii) monotonically increasing in s, or (iii) one switch from monotonically decreasing to monotonically increasing with a minimum at $s_{\min} = \arg\min_{s\in\mathcal{S}} P[Y(s) \leq y|S = t]$. The former two cases can be subsumed in the third if $s_{\min} = s_{(\min)}$ and $s_{\min} = s_{(\max)}$, respectively, where $s_{(\min)}$ denotes the smallest point and $s_{(\max)}$ the largest point in $\mathcal{S}$ (the existence of these points is ensured by the assumption that S is measured on the interval scale). Define $F(y) = P(Y \leq y)$, $F(y|t) = P(Y \leq y|S = t)$, $\mathcal{S}^l_{s,\min} = \{s \in \mathcal{S} : s \leq s_{\min}\}$, and $\mathcal{S}^u_{s,\min} = \{s \in \mathcal{S} : s \geq s_{\min}\}$. Then assumption CXTR yields the following sharp bounds:

Proposition 3. *Let assumption CXTR hold. If $s = t$, then $P[Y(s) \leq y|S = t] = F(y|t)$. If $t < s_{\min}$, then*

$$\begin{array}{lll} s \in \mathcal{S}^l_{s,\min} & s < t \;\Rightarrow & F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) + \dfrac{t - s}{t - s_{(\min)}}[1 - F(y|t)] \\ & s > t \;\Rightarrow & 0 \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) \\ s \notin \mathcal{S}^l_{s,\min} & \;\Rightarrow & 0 \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1 - \dfrac{s_{(\max)} - s}{s_{(\max)} - s_{\min}}[1 - F(y|t)] \end{array}$$

If $t = s_{\min}$, then

$$\begin{array}{ll} s < t \;\Rightarrow & F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) + \dfrac{s_{\min} - s}{s_{\min} - s_{(\min)}}[1 - F(y|t)] \\ s > t \;\Rightarrow & F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1 - \dfrac{s_{(\max)} - s}{s_{(\max)} - s_{\min}}[1 - F(y|t)] \end{array}$$

If $t > s_{\min}$, then

$$\begin{array}{lll} s \in \mathcal{S}^u_{s,\min} & s < t \;\Rightarrow & 0 \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) \\ & s > t \;\Rightarrow & F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1 - \dfrac{s_{(\max)} - s}{s_{(\max)} - t}[1 - F(y|t)] \\ s \notin \mathcal{S}^u_{s,\min} & \;\Rightarrow & 0 \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) + \dfrac{s_{\min} - s}{s_{\min} - s_{(\min)}}[1 - F(y|t)] \end{array}$$

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

Proof. See the appendix.

While the CXTS assumption yields bounds on $P[Y(s) \leq y|S = t]$ over realized states t, for the hypothetical state s fixed, the CXTR assumption yields bounds over hypothetical states s, with the realized state $S = t$ fixed. Thus, the two assumptions make different contributions to the (partial) identification of the potential outcome distribution. Under the convex treatment response assumption, the bounds have the following form:

Corollary 3. *Let assumption CXTR hold. Then,*

$$
\begin{aligned}
s < s_{\min} \;\; \Rightarrow \;\; & P(Y \leq y|s \leq S \leq s_{\min})P(s \leq S \leq s_{\min}) \;\leq\; P[Y(s) \leq y] \;\leq \\
& F(y) + \sum_{s<t<s_{\min}} \left(\frac{t-s}{t-s_{(\min)}}[1-F(y|t)] \right) P(S=t) \\
& + \sum_{t \geq s_{\min}} \left(\frac{s_{\min}-s}{s_{\min}-s_{(\min)}}[1-F(y|t)] \right) P(S=t) \\
s = s_{\min} \;\; \Rightarrow \;\; & F(y|s)P(S=s) \;\leq\; P[Y(s) \leq y] \;\leq\; F(y) \\
s > s_{\min} \;\; \Rightarrow \;\; & P(Y \leq y|s_{\min} \leq S \leq s)P(s_{\min} \leq S \leq s) \;\leq\; P[Y(s) \leq y] \;\leq \\
& P(Y \leq y|S \geq s)P(S \geq s) + P(S \leq s) \\
& - \sum_{t \leq s_{\min}} \left(\frac{s_{(\max)}-s}{s_{(\max)}-s_{\min}}[1-F(y|t)] \right) P(S=t) \\
& - \sum_{s_{\min}<t<s} \left(\frac{s_{(\max)}-s}{s_{(\max)}-t}[1-F(y|t)] \right) P(S=t)
\end{aligned}
$$

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

Proof. See the appendix.

A concave treatment response (CVTR) assumption can be formulated accordingly:

Assumption (CVTR)**.** *For each $y \in \mathcal{Y}$, $(s_1, s_2, t) \in \mathcal{S}^3$, with $s_1 \leq s_2$, and for all $\beta \in [0,1]$, let $P[Y(\beta s_1 + (1-\beta)s_2) \leq y|S=t] \;\geq\; \beta P[Y(s_1) \leq y|S=t] + (1-\beta)P[Y(s_2) \leq y|S=t]$.*

Define $s_{\max} = \arg\max_{s \in \mathcal{S}} P[Y(s) \leq y|S = t]$, $\mathcal{S}^l_{s,\max} = \{s \in \mathcal{S} : s \leq s_{\max}\}$, and $\mathcal{S}^u_{s,\max} = \{s \in \mathcal{S} : s \geq s_{\max}\}$. Following analogous arguments as under the convexity assumption, and the convex/concave treatment selection assumptions, one can derive the following bounds on the counterfactual distributions:

Proposition 4. *Let assumption CVTR hold. If $s = t$, then $P[Y(s) \leq y|S = t] = F(y|t)$.*

If $t < s_{\max}$, then

$$\begin{array}{lll}
s \in \mathcal{S}^l_{s,\max} & s < t \;\Rightarrow & \dfrac{s - s_{(\min)}}{t - s_{(\min)}} F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) \\
 & s > t \;\Rightarrow & F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1 \\
s \notin \mathcal{S}^l_{s,\max} & \;\Rightarrow & \dfrac{s_{(\max)} - s}{s_{(\max)} - s_{\max}} F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1
\end{array}$$

If $t = s_{\max}$, then

$$\begin{array}{ll}
s < t \;\Rightarrow & \dfrac{s - s_{(\min)}}{s_{\max} - s_{(\min)}} F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) \\
s > t \;\Rightarrow & \dfrac{s_{(\max)} - s}{s_{(\max)} - s_{\max}} F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t)
\end{array}$$

If $t > s_{\max}$, then

$$\begin{array}{lll}
s \in \mathcal{S}^u_{s,\max} & s < t \;\Rightarrow & F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1 \\
 & s > t \;\Rightarrow & \dfrac{s_{(\max)} - s}{s_{(\max)} - t} F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; F(y|t) \\
s \notin \mathcal{S}^u_{s,\max} & \;\Rightarrow & \dfrac{s - s_{(\min)}}{s_{\max} - s_{(\min)}} F(y|t) \;\leq\; P[Y(s) \leq y|S = t] \;\leq\; 1
\end{array}$$

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

Proof. See the appendix.

The concave treatment response assumptions places bounds on each counterfactual distribution that in turn can be used to bound the potential outcome distributions:

Corollary 4. *Let assumption CVTR hold. Then,*

$$\begin{array}{ll}
s < s_{\max} \;\Rightarrow & P(Y \leq y|S \leq s)P(S \leq s) + \displaystyle\sum_{s<t<s_{\max}} \left(\frac{s - s_{(\min)}}{t - s_{(\min)}} F(y|t) \right) P(S = t) \\
 & + \displaystyle\sum_{t \geq s_{\max}} \left(\frac{s - s_{(\min)}}{s_{\max} - s_{(\min)}} F(y|t) \right) P(S = t) \;\leq\; P[Y(s) \leq y] \;\leq \\
 & P(Y \leq y|s \leq S \leq s_{\max})P(s \leq S \leq s_{\max}) + 1 - P(s \leq S \leq s_{\max}) \\
s = s_{\max} \;\Rightarrow & F(y) \;\leq\; P[Y(s) \leq y] \;\leq\; F(y|s)P(S = s) + P(S \neq s) \\
s > s_{\max} \;\Rightarrow & P(Y \leq y|S \geq s)P(S \geq s) + \displaystyle\sum_{t \leq s_{\max}} \left(\frac{s_{(\max)} - s}{s_{(\max)} - s_{\max}} F(y|t) \right) P(S = t)
\end{array}$$

$$+\sum_{s_{\max}<t<s}\left(\frac{s_{(\max)}-s}{s_{(\max)}-t}F(y|t)\right)P(S=t) \ \leq \ P[Y(s)\leq y] \ \leq$$
$$P(Y\leq y|s_{\max}\leq S\leq s)P(s_{\max}\leq S\leq s)+1-P(s_{\max}\leq S\leq s)$$

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

Proof. See the appendix.

While each assumption alone is not refutable, a joint set of assumptions is refutable using the empirical evidence alone. For example, if the convex treatment selection assumption is combined with the convex treatment response assumption, then the observed distribution $P(Y \leq y|S)$ is convex in S. Thus, if a convex shape of $P(Y \leq y|S)$ is rejected by the data, then the joint assumptions CXTS/CXTR can be rejected as well. A particularly appealing case is obtained if $s_{\min} = t_{\min}$ which then becomes a global minimum point. Analogous arguments hold for the combination of concavity assumptions.

3 Bounds on the Effect of Education on Smoking

The causal effect of education and smoking is used to illustrate the convexity assumptions and the construction of bounds. Several explanations exist in favor of a causal mechanism, including theories of productive efficiency (Grossman 1972) and allocative efficiency (Kenkel 1991; Rosenzweig 1995). A competing interpretation asserts the existence of individual time preferences as unobserved background factor that generates a negative correlation even in the absence of a causal effect (Farrell and Fuchs 1982).

More recently, Currie and Moretti (2003), Kenkel *et al.* (2006), de Walque (2007), Grimard and Parent (2007,) Gilman *et al.* (2008), and Tenn *et al.* (2008) employed instrumental variables to estimate the effect of education on smoking. These studies are not conclusive, however, regarding the magnitude of the impact, ranging from negligibly small to negative and significant. While substantial effort is made to justify the instruments in each of these papers, their validity is not entirely without debate (see Tenn *et al.* 2008).

The analysis here adds to the previous literature by investigating the effect of education on different parts of the smoking distribution. I will make use of the convexity assumptions outlined above and thus will not require exogenous instruments or exclusion restrictions. 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 40 and 55, and education levels of 6 to 16 years of schooling, which gives a total of 1,180 observations. Table 1 and Figure 1 display features of the smoking distribution conditional on different levels of schooling. Consider Table 1 first. The schooling distribution indicates that the largest fraction of people in the sample obtained a high school degree, followed by those who obtained vocational, and undergraduate college or university education.

— Insert Table 1 and Figure 1 about here —

Table 1 also reports the probability of non-smoking, and smoking no more than 10 (20) cigarettes on average per day, conditional on schooling. The shape of the smoking distribution, as a function of schooling, at each of these thresholds is convex over the most part of the schooling support (as shown in Figure 1). Exceptions are the upper right two points for the probability of smoking less than 20 cigarettes conditional on 15/16 years of schooling. However, there is only a small number of observations in this category, and the corresponding confidence intervals are wide enough to be consistent with an everywhere convex function.

A consistent minimum is obtained at about eight years of schooling. Following the above arguments, I will maintain the assumption that $t_{\min} = \arg\min_{t \in \mathcal{S}} P(Y(s) \leq y | S = t)$ and $s_{\min} = t_{\min} = 8$. Figure 2 shows the bounds on the potential outcome distribution using the empirical evidence alone (the no-assumptions bounds), under the convex treatment selection (CXTS) assumption only, under the convex treatment response (CXTR) assumption only, and if the latter two types are imposed jointly (CXTS/CXTR).

— Insert Figure 2 about here —

The CXTS and the CXTR assumptions contribute differently to the identification of the counterfactual distributions, and hence, the width of the bounds on the potential outcome distributions differs between these two types of assumptions. Moreover, the joint assumptions yield bounds that are even tighter than the intersection bounds because CXTR and CXTS can make both contributions to more informative upper and/or more informative lower bounds. Figure 2 shows that there is a substantial improvement in the bounds by using the convexity assumptions. The results indicate, for example, that the probability of non-smoking lies between about 41 and 51 percent for those with 8 years of schooling, compared to about 2 and 85 percent using the no-assumptions bounds.

It should be noted that the convexity assumptions alone do not suffice to evaluate whether education has a positive, a negative, or no effect at all on the probability to smoke (nor on the distribution at the other thresholds shown in Figure 2). This follows because the intersection of the bounds for all schooling levels is a non-empty set. Thus, without imposing additional restrictions on the data, the analysis presented here is consistent with the notion that the association between education and smoking is merely related to unobserved background factors, even in the absence of a causal mechanism.

Figure 3 provides an additional set of bounds using assumptions of semi-monotonicity (Boes 2009). The minimum monotone treatment selection (MMTS) and minimum monotone treatment response (MMTR) assumptions relax the assumptions of convexity to monotonically decreasing (increasing) $P[Y(s) \leq y|S = t]$ in t and s to the left (right) of the conjectured minimum point. The results indicate that there is a moderate information gain by imposing convexity as opposed to semi-monotonicity.

— Insert Figure 3 about here —

4 Concluding Remarks

This paper proposed two weak convexity (or concavity) assumptions on the counterfactual distributions and derives their partial identification results for the potential outcome distributions. First, the convex treatment selection assumption, which imposes constraints on the variation of the potential outcome distribution conditional on realized states. Second, the convex treatment response assumption, which restricts the shape of the outcome distribution over potential states. Both types of assumptions jointly are refutable by the observed data distribution, and they may significantly improve over the no-assumptions bounds.

The bounding strategy is limited in two ways. First, the data requirements are vast because the outcome distribution in each state is estimated separately. Furthermore, without additional assumptions it is generally impossible to extrapolate to outcomes not actually observed. Second, 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, estimation of bounds for continuous outcomes is generally more challenging than for discrete outcomes as in the example above. Third, the uncertainty in the support of the convexity/concavity assumption and the location of the extremum points must be accounted for in order to derive confidence intervals with a pre-defined coverage probability.

References

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), Bounds on Counterfactual Distributions Under Semi-Monotonicity Constraints, *SOI Working Paper* 0920.

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.

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.

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

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.

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.

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.

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.

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 (2009), Concave-Monotone Treatment Response and Monotone Treatment Selection: With an Application to the Returns to Schooling, available online at http://ssrn.com/abstract=1512717.

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

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.

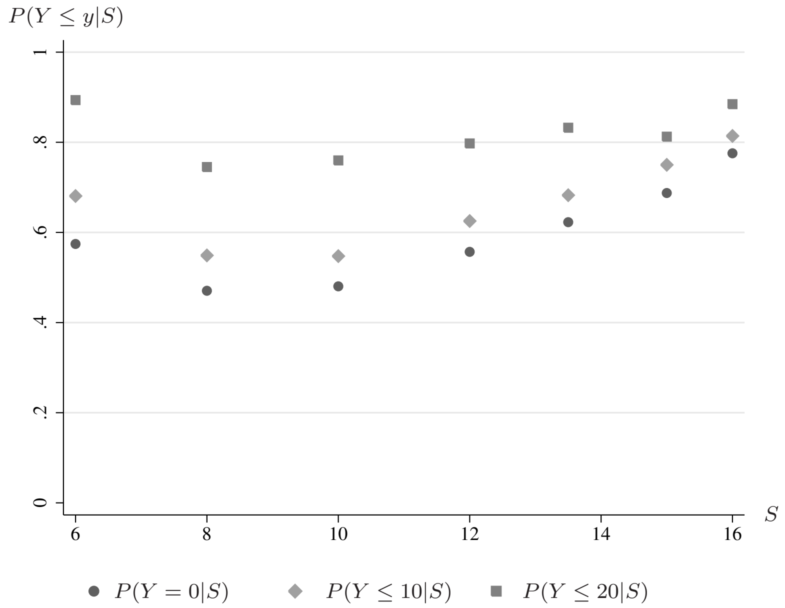
Tables and Figures

Table 1: Smoking distribution conditional on schooling

Years of schooling	$P(S)$	$P(Y=0\|S)$	$P(Y\leq 10\|S)$	$P(Y\leq 20\|S)$
6	0.040 [47]	0.575	0.681	0.894
8	0.043 [51]	0.471	0.549	0.745
10	0.152 [179]	0.480	0.548	0.760
12	0.335 [395]	0.557	0.625	0.798
13.5	0.142 [167]	0.623	0.683	0.832
15	0.027 [32]	0.688	0.750	0.813
16	0.132 [156]	0.776	0.814	0.884

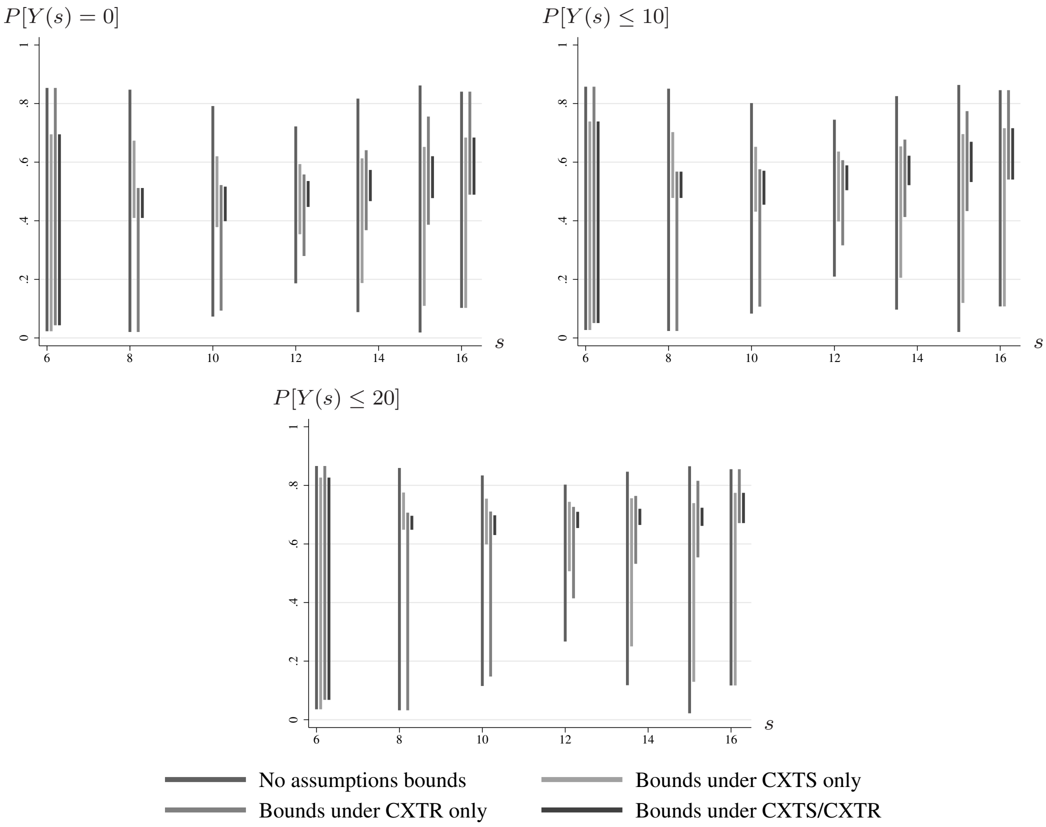
Source: Smoking Supplement of the 1979 US National Health Interview Survey, own calculations. *Notes:* The estimates are based on a random sample of 1,180 observations on employed white men aged 40-55. The numbers of observations used to estimate each cumulative distribution are reported in square brackets. Y denotes the average number of cigarettes smoked per day, S denotes the years of schooling.

Figure 1: Convex shape of observed smoking distribution by schooling



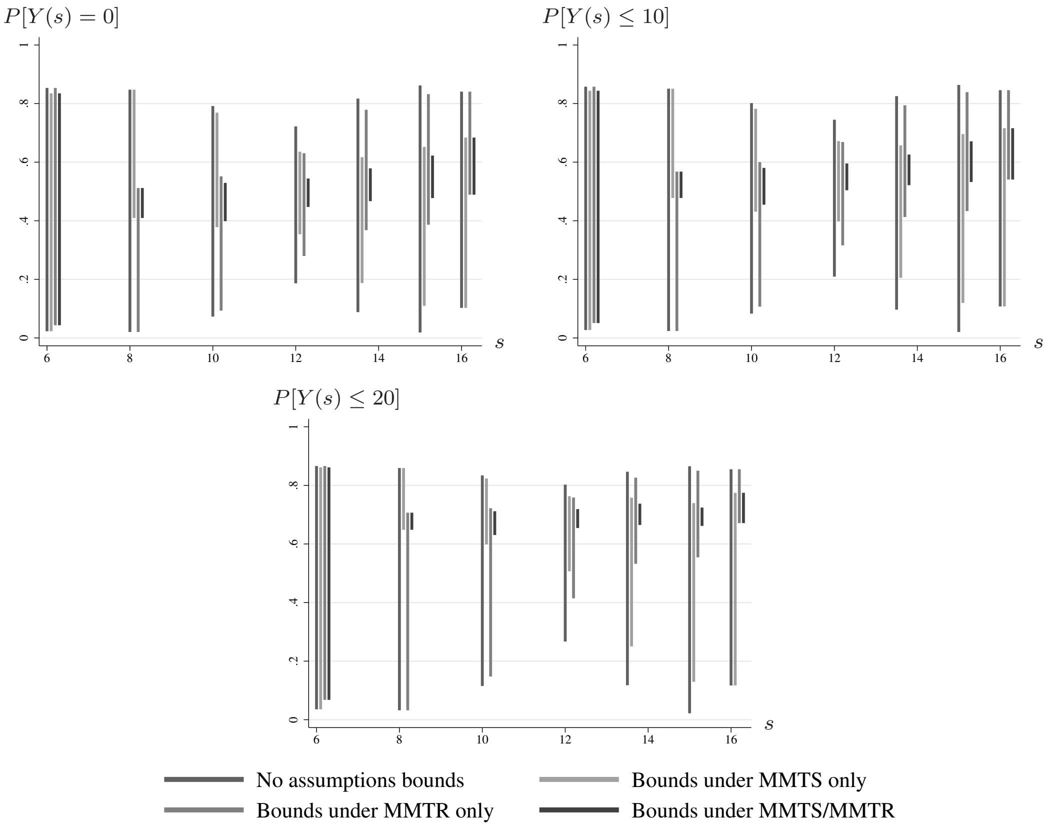
Source: Smoking Supplement of the 1979 US National Health Interview Survey, own calculations. *Notes:* The estimates are based on a random sample of 1,180 observations on employed white men aged 40-55. Y is the average number of cigarettes smoked per day, S is the years of schooling.

Figure 2: Convexity bounds on the potential smoking distribution by schooling



Source: Smoking Supplement of the 1979 US National Health Interview Survey, own calculations. *Notes:* The estimates are based on a random sample of 1,103 observations on employed white men aged 35-60. Left bars are obtained using the empirical evidence alone. The CXTS assumption asserts convexity in u with minimum point $2 = \arg\min_{u \in \mathcal{S}} P[Y(s) \leq y | S = u]$ (middle left bars), the CXTR assumption asserts convexity in s with minimum point $2 = \arg\min_{s \in \mathcal{S}} P[Y(s) \leq y | S = u]$ (middle right bars). The right bars indicate the bounds obtained under the joint set of assumptions CXTS and CXTR.

Figure 3: Monotonicity bounds on the potential smoking distribution by schooling



Source: Smoking Supplement of the 1979 US National Health Interview Survey, own calculations. *Notes:* The estimates are based on a random sample of 1,180 observations on employed white men aged 40-55. Left bars are obtained using the empirical evidence alone. The MMTS assumption asserts monotonicity in u around the minimum at $2 = \arg\min_{u \in \mathcal{S}} P[Y(s) \leq y | S = u]$ (middle left bars), the MMTR assumption asserts monotonicity in s around the minimum at $2 = \arg\min_{s \in \mathcal{S}} P[Y(s) \leq y | S = u]$ (middle right bars). The right bars indicate the bounds obtained under the joint set of assumptions MMTS and MMTR.

Appendix: Proofs

Proof of Proposition 1.

Consider the case $s < t_{\min}$ (the case $s > t_{\min}$ is symmetric, the case $s = t_{\min}$ immediately follows). If $t < s$, then $F(y|s) = P(Y \leq y|S = s)$ is a lower bound of $P[Y(s) \leq y|S = t]$ by monotonicity (recall that $P[Y(s) \leq y|S = t]$ is monotonically decreasing in S for all $t \in \mathcal{S}^l_{t,\min}$). The upper bound is given by $\alpha^* P[Y(s) \leq y|S = t_1] + (1-\alpha^*)P[Y(s) \leq y|S = t_2]$ (which follows from convexity). Note that $\alpha^* = (t_2 - t)/(t_2 - t_1)$, i.e., α^* is monotonically increasing in t_2 and t_1. The smallest upper bound of $P[Y(s) \leq y|S = t_1]$ is one since $t_1 \leq t$, the smallest upper bound of $P[Y(s) \leq y|S = t_2]$ is $F(y|s)$ for all t_2 in the interval $[s, t_{\min}]$. Thus, the upper bound of $P[Y(s) \leq y|S = t]$ can be written as

$$P[Y(s) \leq y|S = t] \leq \alpha^* + (1 - \alpha^*)F(y|s) = F(y|s) + \alpha^*[1 - F(y|s)]$$

Since the upper bound holds for all $t_1 \leq t$ and all $t_2 \in [s, t_{\min}]$, and the upper bound is smaller the smaller α^*, the optimal (α^*-minimizing) choice of t_1, t_2 is $t_1 = t_{(\min)}$ and $t_2 = s$ such that $\alpha^* = (s - t)/(s - t_{(\min)})$. If $t = s$, then $P[Y(s) \leq y|S = t] = F(y|s)$. If $s < t \leq t_{\min}$, then $P[Y(s) \leq y|S = t]$ is bounded from below by zero (by monotonicity). The upper bound is given by $\alpha^* P[Y(s) \leq y|S = t_1] + (1 - \alpha^*)P[Y(s) \leq y|S = t_2]$ (by convexity). The smallest upper bounds on $P[Y(s) \leq y|S = t_1]$ and $P[Y(s) \leq y|S = t_2]$ is $F(y|s)$ which therefore is the upper bound of $P[Y(s) \leq y|S = t]$. If $t > t_{\min}$, then $P[Y(s) \leq y|S = t]$ is bounded from below by zero (by monotonicity). Using the convexity argument, the smallest upper bound of $P[Y(s) \leq y|S = t_1]$ is given by $F(y|s)$ (in the interval $[s, t_{\min}]$), the smallest upper bound of $P[Y(s) \leq y|S = t_2]$ is one. Thus, the upper bound of $P[Y(s) \leq y|S = t]$ can be written as

$$P[Y(s) \leq y|S = t] \leq \alpha^* F(y|s) + (1 - \alpha^*) = 1 - \alpha^*[1 - F(y|s)]$$

Since the upper bound holds for all $t_1 \in [s, t_{\min}]$ and all $t_2 \geq t$, and the upper bound is smaller the larger α^*, the optimal (α^*-maximizing) choice of t_1, t_2 is $t_1 = t_{\min}$ and $t_2 = t_{(\max)}$ such that $\alpha^* = (t_{(\max)} - t)/(t_{(\max)} - t_{\min})$. Since assumption CXTS is the only assumption invoked on the data, each lower and upper bound is sharp. Hence, the overall bounds are sharp. □

Proof of Corollary 1.

Consider the case $s < t_{\min}$. The potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] &= P(Y \leq y|S = s)P(S = s) + \sum_{t<s} P[Y(s) \leq y|S = t]P(S = t) \\ &+ \sum_{s<t\leq t_{\min}} P[Y(s) \leq y|S = t]P(S = t) \\ &+ \sum_{t>t_{\min}} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

Proposition 1 provides sharp bounds on each counterfactual distribution. For the first sum, the lower and upper bounds are $F(y|s)$ and $F(y|s)+(s-t)/(s-t_{(\min)})[1-F(y|s)]$, respectively. For the second sum, the bounds are zero and $F(y|s)$. For the third sum, the bounds are zero and $1 - (t_{(\max)} - t)/(t_{(\max)} - t_{\min})[1 - F(y|s)]$, which yields the stated bounds.

For $s = t_{\min}$ the potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] &= P(Y \leq y|S = s)P(S = s) + \sum_{t<s} P[Y(s) \leq y|S = t]P(S = t) \\ &+ \sum_{t<s} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

Using the results of Proposition 1, the lower bound of $P[Y(s) \leq y|S = t]$ for all $t \neq s$ is $F(y|s)$. The upper bound is $F(y|s) + (t_{\min} - t)/(t_{\min} - t_{(\min)})[1 - F(y|s)]$ if $t < s$, the upper bound is $1 - (t_{(\max)} - t)/(t_{(\max)} - t_{\min})[1 - F(y|s)]$ if $t > s$.

Finally, consider the case $s > t_{\min}$. The potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] &= P(Y \leq y|S = s)P(S = s) + \sum_{t<t_{\min}} P[Y(s) \leq y|S = t]P(S = t) \\ &+ \sum_{t_{\min}\leq t<s} P[Y(s) \leq y|S = t]P(S = t) \\ &+ \sum_{t>s} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

For the first sum, the bounds as stated in Proposition 1 are zero and $F(y|s)+(t_{\min}-t)/(t_{\min}-t_{(\min)})[1 - F(y|s)]$, respectively. For the second sum, the bounds are zero and $F(y|s)$. For the third sum, the bounds are $F(y|s)$ and $1 - (t_{(\max)} - t)/(t_{(\max)} - s)[1 - F(y|s)]$.

Since the potential outcome distribution is a sum of counterfactual distributions, and the bounds on the latter are sharp by Proposition 1, the overall bounds are sharp. □

Proof of Proposition 2.

Consider the case $s < t_{\max}$ (the case $s > t_{\max}$ is symmetric, the case $s = t_{\max}$ immediately follows). If $t < s$, then $F(y|s) = P(Y \leq y|S = s)$ is an upper bound of $P[Y(s) \leq y|S = t]$ by monotonicity (recall that $P[Y(s) \leq y|S = t]$ is monotonically increasing in S for all $t \in \mathcal{S}^l_{t,\max}$). The lower bound is given by $\alpha^* P[Y(s) \leq y|S = t_1] + (1-\alpha^*)P[Y(s) \leq y|S = t_2]$ (which follows from concavity). The largest lower bound of $P[Y(s) \leq y|S = t_1]$ is zero since $t_1 \leq t$, the largest lower bound of $P[Y(s) \leq y|S = t_2]$ is $F(y|s)$ for all t_2 in the interval $[s, t_{\max}]$. Thus, the lower bound of $P[Y(s) \leq y|S = t]$ can be written as

$$P[Y(s) \leq y|S = t] \geq (1 - \alpha^*)F(y|s)$$

Since the lower bound holds for all $t_1 \leq t$ and all $t_2 \in [s, t_{\max}]$, and the lower bound is larger the larger $1 - \alpha^*$, or the smaller α^*, the optimal (α^*-minimizing) choice of t_1, t_2 is $t_1 = t_{(\min)}$ and $t_2 = s$ such that $1 - \alpha^* = (t - t_{(\min)})/(s - t_{(\min)})$. If $t = s$, then $P[Y(s) \leq y|S = t] = F(y|s)$. If $s < t \leq t_{\max}$, then $P[Y(s) \leq y|S = t]$ is bounded from above by one (by monotonicity). The lower bound is given by $\alpha^* P[Y(s) \leq y|S = t_1] + (1 - \alpha^*)P[Y(s) \leq y|S = t_2]$ (by concavity). The largest lower bounds on $P[Y(s) \leq y|S = t_1]$ and $P[Y(s) \leq y|S = t_2]$ is $F(y|s)$ which therefore is the lower bound of $P[Y(s) \leq y|S = t]$. If $t > t_{\max}$, then $P[Y(s) \leq y|S = t]$ is bounded from above by one (by monotonicity). Using the concavity argument, the largest lower bound of $P[Y(s) \leq y|S = t_1]$ is given by $F(y|s)$ (in the interval $[s, t_{\max}]$), the largest lower bound of $P[Y(s) \leq y|S = t_2]$ is zero. Thus, the lower bound of $P[Y(s) \leq y|S = t]$ can be written as

$$P[Y(s) \leq y|S = t] \geq \alpha^* F(y|s)$$

Since the lower bound holds for all $t_1 \in [s, t_{\max}]$ and all $t_2 \geq t$, and the lower bound is larger the larger α^*, the optimal (α^*-maximizing) choice of t_1, t_2 is $t_1 = t_{\max}$ and $t_2 = t_{(\max)}$ such that $\alpha^* = (t_{(\max)} - t)/(t_{(\max)} - t_{\max})$. Since assumption CVTS is the only assumption invoked on the data, each lower and upper bound is sharp. Hence, the overall bounds are sharp. □

Proof of Corollary 2.

Consider the case $s < t_{\max}$. The potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] \quad &= \quad P(Y \leq y|S = s)P(S = s) + \sum_{t<s} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + \sum_{s<t\leq t_{\max}} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

$$+\sum_{t>t_{\max}} P[Y(s)\leq y|S=t]P(S=t)$$

Proposition 2 provides sharp bounds on each counterfactual distribution. For the first sum, the lower and upper bounds are $(t-t_{(\min)})/(s-t_{(\min)})F(y|s)$ and $F(y|s)$, respectively. For the second sum, the bounds are $F(y|s)$ and one. For the third sum, the bounds are $(t_{(\max)}-t)/(t_{(\max)}-t_{\max})[1-F(y|s)]$ and one, which yields the stated bounds.

For $s=t_{\max}$ the potential outcome distribution can be written as

$$\begin{aligned}P[Y(s)\leq y] &= P(Y\leq y|S=s)P(S=s)+\sum_{t<s} P[Y(s)\leq y|S=t]P(S=t)\\ &\quad +\sum_{t<s} P[Y(s)\leq y|S=t]P(S=t)\end{aligned}$$

Using the results of Proposition 2, the upper bound of $P[Y(s)\leq y|S=t]$ for all $t\neq s$ is $F(y|s)$. The lower bound is $(t-t_{(\min)})/(t_{\max}-t_{(\min)})F(y|s)$ if $t<s$, the lower bound is $(t_{(\max)}-t)/(t_{(\max)}-t_{\max})F(y|s)$ if $t>s$.

Finally, consider the case $s>t_{\max}$. The potential outcome distribution can be written as

$$\begin{aligned}P[Y(s)\leq y] &= P(Y\leq y|S=s)P(S=s)+\sum_{t<t_{\max}} P[Y(s)\leq y|S=t]P(S=t)\\ &\quad +\sum_{t_{\max}\leq t<s} P[Y(s)\leq y|S=t]P(S=t)\\ &\quad +\sum_{t>s} P[Y(s)\leq y|S=t]P(S=t)\end{aligned}$$

For the first sum, the bounds as stated in Proposition 2 are $(t-t_{(\min)})/(t_{\max}-t_{(\min)})F(y|s)$ and one, respectively. For the second sum, the bounds are $F(y|s)$ and one. For the third sum, the bounds are $(t_{(\max)}-t)/(t_{(\max)}-s)F(y|s)$ and $F(y|s)$.

Since the potential outcome distribution is a sum of counterfactual distributions, and the bounds on the latter are sharp by Proposition 2, the overall bounds are sharp. □

Proof of Proposition 3.

Consider the case $t<s_{\min}$ (the case $t>s_{\min}$ is symmetric, the case $t=s_{\min}$ immediately follows). If $s<t$, then $F(y|t)=P(Y\leq y|S=t)$ is a lower bound of $P[Y(s)\leq y|S=t]$ by monotonicity (recall that $P[Y(s)\leq y|S=t]$ is monotonically decreasing in s for all $s\in\mathcal{S}^{l}_{s,\min}$). The upper bound is given by $\beta^{*}P[Y(s_1)\leq y|S=t]+(1-\beta^{*})P[Y(s_2)\leq y|S=t]$ (which follows from convexity). Note that $\beta^{*}=(s_2-s)/(s_2-s_1)$, i.e., β^{*} is monotonically increasing in s_2 and s_1. The smallest upper bound of $P[Y(s_1)\leq y|S=t]$ is one since $s_1\leq s$, the smallest

upper bound of $P[Y(s_2) \leq y|S=t]$ is $F(y|t)$ for all s_2 in the interval $[t, s_{\min}]$. Thus, the upper bound of $P[Y(s) \leq y|S=t]$ can be written as

$$P[Y(s) \leq y|S=t] \leq \beta^* + (1-\beta^*)F(y|t) = F(y|t) + \beta^*[1-F(y|t)]$$

Since the upper bound holds for all $s_1 \leq s$ and all $s_2 \in [t, s_{\min}]$, and the upper bound is smaller the smaller β^*, the optimal (β^*-minimizing) choice of s_1, s_2 is $s_1 = s_{(\min)}$ and $s_2 = t$ such that $\beta^* = (t-s)/(t-s_{(\min)})$. If $s = t$, then $P[Y(s) \leq y|S=t] = F(y|t)$. If $t < s \leq s_{\min}$, then $P[Y(s) \leq y|S=t]$ is bounded from below by zero (by monotonicity). The upper bound is given by $\beta^* P[Y(s_1) \leq y|S=t] + (1-\beta^*)P[Y(s_2) \leq y|S=t]$ (by convexity). The smallest upper bounds on $P[Y(s_1) \leq y|S=t]$ and $P[Y(s_2) \leq y|S=t]$ is $F(y|t)$ which therefore is the upper bound of $P[Y(s) \leq y|S=t]$. If $s > s_{\min}$, then $P[Y(s) \leq y|S=t]$ is bounded from below by zero (by monotonicity). Using the convexity argument, the smallest upper bound of $P[Y(s_1) \leq y|S=t]$ is given by $F(y|t)$ (in the interval $[t, s_{\min}]$), the smallest upper bound of $P[Y(s_2) \leq y|S=t]$ is one. Thus, the upper bound of $P[Y(s) \leq y|S=t]$ can be written as

$$P[Y(s) \leq y|S=t] \leq \beta^* F(y|t) + (1-\beta^*) = 1 - \beta^*[1-F(y|t)]$$

Since the upper bound holds for all $s_1 \in [t, s_{\min}]$ and all $s_2 \geq s$, and the upper bound is smaller the larger β^*, the optimal (β^*-maximizing) choice of s_1, s_2 is $s_1 = s_{\min}$ and $s_2 = s_{(\max)}$ such that $\beta^* = (s_{(\max)} - s)/(s_{(\max)} - s_{\min})$. Since assumption CXTR is the only assumption invoked on the data, each lower and upper bound is sharp. Hence, the overall bounds are sharp. $\square$

Proof of Corollary 3.

Consider the case $s < s_{\min}$. The potential outcome distribution can be written as

$$\begin{aligned}
P[Y(s) \leq y] \quad = \quad & P(Y \leq y|S=s)P(S=s) + \sum_{t<s} P[Y(s) \leq y|S=t]P(S=t) \\
& + \sum_{s<t<s_{\min}} P[Y(s) \leq y|S=t]P(S=t) \\
& + P[Y(s) \leq y|S=s_{\min}]P(S=s_{\min}) \\
& + \sum_{t>s_{\min}} P[Y(s) \leq y|S=t]P(S=t)
\end{aligned}$$

Proposition 3 provides sharp bounds on each counterfactual distribution. For the first sum, the lower and upper bounds are zero and $F(y|t)$, respectively. For the second sum, the bounds are $F(y|t)$ and $F(y|t) + (t-s)/(t-s_{\min})[1-F(y|t)]$. The counterfactual distributions in the third term and the last sum is bounded by zero and $F(y|t) + (s_{\min} - s)/(s_{\min} - s_{(\min)})[1 - F(y|t)]$. Summarizing terms yields the stated bounds.

For $s = s_{\min}$ the potential outcome distribution can be written as

$$\begin{aligned} P[Y(s_{\min}) \leq y] &= P(Y \leq y|S = s_{\min})P(S = s_{\min}) \\ &\quad + \sum_{t<s_{\min}} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + \sum_{t>s_{\min}} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

Using the results of Proposition 3, the lower bound of $P[Y(s) \leq y|S = t]$ for all $t < s_{\min}$ and for all $t > s_{\min}$ is zero, the upper bound is $F(y|t)$.

Finally, consider the case $s > s_{\min}$. The potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] &= P(Y \leq y|S = s)P(S = s) + \sum_{t<s_{\min}} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + P[Y(s) \leq y|S = s_{\min}]P(S = s_{\min}) \\ &\quad + \sum_{s_{\min}<t<s} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + \sum_{t>s} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

For the first sum, the bounds are zero and $1-(s_{(\max)}-s)/(s_{(\max)}-s_{\min})[1-F(y|t)]$, respectively, as stated in Proposition 3. The counterfactial distributions in the third term and the second sum are bounded by $F(y|t)$ and $1 - (s_{(\max)} - s)/(s_{(\max)} - t)[1 - F(y|t)]$. The bounds in the last sum are zero and $F(y|t)]$.

Since the potential outcome distribution is a sum of counterfactual distributions, and the bounds on the latter are sharp by Proposition 3, the overall bounds are sharp. □

Proof of Proposition 4.

Consider the case $t < s_{\max}$ (the case $t > s_{\max}$ is symmetric, the case $t = s_{\max}$ immediately follows). If $s < t$, then $F(y|t) = P(Y \leq y|S = t)$ is an upper bound of $P[Y(s) \leq y|S = t]$ by monotonicity (recall that $P[Y(s) \leq y|S = t]$ is monotonically increasing in s for all $s \in \mathcal{S}^l_{s,\max}$). The lower bound is given by $\beta^* P[Y(s_1) \leq y|S = t] + (1-\beta^*)P[Y(s_2) \leq y|S = t]$ (which follows from concavity). Note that $\beta^* = (s_2 - s)/(s_2 - s_1)$, i.e., β^* is monotonically increasing in s_2 and s_1. The largest lower bound of $P[Y(s_1) \leq y|S = t]$ is zero since $s_1 \leq s$, the largest lower bound of $P[Y(s_2) \leq y|S = t]$ is $F(y|t)$ for all s_2 in the interval $[t, s_{\max}]$. Thus, the lower bound of $P[Y(s) \leq y|S = t]$ can be written as

$$P[Y(s) \leq y|S = t] \geq (1 - \beta^*)F(y|t)$$

Since the lower bound holds for all $s_1 \leq s$ and all $s_2 \in [t, s_{\max}]$, and the lower bound is larger the smaller β^*, the optimal (β^*-minimizing) choice of s_1, s_2 is $s_1 = s_{(\min)}$ and $s_2 = t$ such that $1 - \beta^* = (s - s_{(\min)})/(t - s_{(\min)})$. If $s = t$, then $P[Y(s) \leq y|S = t] = F(y|t)$. If $t < s \leq s_{\max}$, then $P[Y(s) \leq y|S = t]$ is bounded from above by one (by monotonicity). The lower bound is given by $\beta^* P[Y(s_1) \leq y|S = t] + (1 - \beta^*)P[Y(s_2) \leq y|S = t]$ (by concavity). The largest lower bounds on $P[Y(s_1) \leq y|S = t]$ and $P[Y(s_2) \leq y|S = t]$ is $F(y|t)$ which therefore is the lower bound of $P[Y(s) \leq y|S = t]$. If $s > s_{\max}$, then $P[Y(s) \leq y|S = t]$ is bounded from above by one (by monotonicity). Using the concavity argument, the largest lower bound of $P[Y(s_1) \leq y|S = t]$ is given by $F(y|t)$ (in the interval $[t, s_{\max}]$), the largest lower bound of $P[Y(s_2) \leq y|S = t]$ is zero. Thus, the lower bound of $P[Y(s) \leq y|S = t]$ can be written as

$$P[Y(s) \leq y|S = t] \geq \beta^* F(y|t)$$

Since the lower bound holds for all $s_1 \in [t, s_{\max}]$ and all $s_2 \geq s$, and the lower bound is larger the larger β^*, the optimal (β^*-maximizing) choice of s_1, s_2 is $s_1 = s_{\max}$ and $s_2 = s_{(\max)}$ such that $\beta^* = (s_{(\max)} - s)/(s_{(\max)} - s_{\max})$. Since assumption CVTR is the only assumption invoked on the data, each lower and upper bound is sharp. Hence, the overall bounds are sharp. $\square$

Proof of Corollary 4.

Consider the case $s < s_{\max}$. The potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] \quad = \quad & P(Y \leq y|S = s)P(S = s) + \sum_{t<s} P[Y(s) \leq y|S = t]P(S = t) \\ & + \sum_{s<t<s_{\max}} P[Y(s) \leq y|S = t]P(S = t) \\ & + P[Y(s) \leq y|S = s_{\max}]P(S = s_{\max}) \\ & + \sum_{t>s_{\max}} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

Proposition 4 provides sharp bounds on each counterfactual distribution. For the first sum, the lower and upper bounds are $F(y|t)$ and one, respectively. For the second sum, the bounds are $(s - s_{(\min)})/(t - s_{(\min)})F(y|t)$ and $F(y|t)$. The counterfactual distributions in the third term and the last sum are bounded by $(s - s_{(\min)})/(s_{\max} - s_{(\min)})F(y|t)$. Summarizing terms yields the stated bounds.

For $s = s_{\max}$ the potential outcome distribution can be written as

$$\begin{aligned} P[Y(s_{\max}) \leq y] &= P(Y \leq y|S = s_{\max})P(S = s_{\max}) \\ &\quad + \sum_{t<s_{\max}} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + \sum_{t>s_{\max}} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

Using the results of Proposition 4, the lower bound of $P[Y(s) \leq y|S = t]$ for all $t < s_{\max}$ and for all $t > s_{\max}$ is $F(y|t)$, the upper bound is one.

Finally, consider the case $s > s_{\max}$. The potential outcome distribution can be written as

$$\begin{aligned} P[Y(s) \leq y] &= P(Y \leq y|S = s)P(S = s) + \sum_{t<s_{\max}} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + P[Y(s) \leq y|S = s_{\max}]P(S = s_{\max}) \\ &\quad + \sum_{s_{\max}<t<s} P[Y(s) \leq y|S = t]P(S = t) \\ &\quad + \sum_{t>s} P[Y(s) \leq y|S = t]P(S = t) \end{aligned}$$

For the first sum, the bounds are $(s_{(\max)} - s)/(s_{(\max)} - s_{\max})F(y|t)$ and one, respectively, as stated in Proposition 4. The counterfactual distributions in the third term and the second sum are bounded by $(s_{(\max)} - s)/(s_{(\max)} - t)F(y|t)$ and $F(y|t)$. The counterfactual distributions in the last sum are bounded by $F(y|t)]$ and one.

Since the potential outcome distribution is a sum of counterfactual distributions, and the bounds on the latter are sharp by Proposition 4, the overall bounds are sharp. □

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.uzh.ch/research/wp_en.html

1001	Convex Treatment Response and Treatment Selection, Stefan Boes, December 2009, 28 p.
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	Economic Well-Being, Social Mobility, and Preferences for Income Redistribution: Evidence from a Discrete Choice Experiment, Ilja Neustadt, Peter Zweifel, July 2009, revised January 2010, 33 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.
0814 Competitive Markets without Commitment, Nick Netzer, Florian Scheuer, November 2008, 65 p.
0813 Scope of Electricity Efficiency Improvement in Switzerland until 2035, Boris Krey, October 2008, 25 p.
0812 Efficient Electricity Portfolios for the United States and Switzerland: An Investor View, Boris Krey, Peter Zweifel, October 2008, 26 p.
0811 A welfare analysis of "junk" information and spam filters; Josef Falkinger, October 2008, 33 p.
0810 Why does the amount of income redistribution differ between United States and Europe? The Janus face of Switzerland; Sule Akkoyunlu, Ilja Neustadt, Peter Zweifel, September 2008, 32 p.
0809 Promoting Renewable Electricity Generation in Imperfect Markets: Price vs. Quantity Policies; Reinhard Madlener, Weiyu Gao, Ilja Neustadt, Peter Zweifel, July 2008, 34p.
0808 Is there a U-shaped Relation between Competition and Investment? Dario Sacco, July 2008, 26p.
0807 Competition and Innovation: An Experimental Investigation, May 2008, 20 p.
0806 All-Pay Auctions with Negative Prize Externalities: Theory and Experimental Evidence, May 2008, 31 p.
0805 Between Agora and Shopping Mall, Josef Falkinger, May 2008, 31 p.
0804 Provision of Public Goods in a Federalist Country: Tiebout Competition, Fiscal Equalization, and Incentives for Efficiency in Switzerland, Philippe Widmer, Peter Zweifel, April 2008, 22 p.
0803 Stochastic Expected Utility and Prospect Theory in a Horse Race: A Finite Mixture Approach, Adrian Bruhin, March 2008, 25 p.
0802 The effect of trade openness on optimal government size under endogenous firm entry, Sandra Hanslin, March 2008, 31 p.
0801 Managed Care Konzepte und Lösungsansätze – Ein internationaler Vergleich aus schweizerischer Sicht, Johannes Schoder, Peter Zweifel, February 2008, 23 p.
0719 Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, Dennis Gärtner, December 2007, 8 p.
0718 Monoplistic Screening under Learning by Doing, Dennis Gärtner, December 2007, 29 p.
0717 An analysis of the Swiss vote on the use of genetically modified crops, Felix Schläpfer, November 2007, 23 p.
0716 The relation between competition and innovation – Why is it such a mess? Armin Schmutzler, November 2007, revised January 2010, 37 p.
0715 Contingent Valuation: A New Perspective, Felix Schläpfer, November 2007, 32 p.
0714 Competition and Innovation: An Experimental Investigation, Dario Sacco, October 2007, 36p.
0713 Hedonic Adaptation to Living Standards and the Hidden Cost of Parental Income, Stefan Boes, Kevin Staub, Rainer Winkelmann, October 2007, 18p.
0712 Competitive Politics, Simplified Heuristics, and Preferences for Public Goods, Felix Schläpfer, Marcel Schmitt, Anna Roschewitz, September 2007, 40p.
0711 Self-Reinforcing Market Dominance, Daniel Halbheer, Ernst Fehr, Lorenz Goette, Armin Schmutzler, August 2007, 34p.