Contents

11
Partial identification and missing data

Identification refers to a description of parameters that can, in principle, be uncovered from arbitrarily large samples.[1] Hence, if parameters are unidentified it is fruitless to engage in estimation and inference. On the other hand, when parameters are identified then it's meaningful to engage in estimation and inference strategies. Much of the extant discussion focuses on point identification but bounds or partial identification can illuminate, say, what conclusions can be drawn from the data alone and how dependent point identification is on conditions that may be of suspect credibility. Point and partial identification are defined by the identification region or feasible values the quantity of interest can take given the sampling process and distributional conditions maintained. A quantity is point identified if the identification region is a single value and partially identified if the identification region involves multiple values but less than the maximum region. Partial identification is described by a region when there is full support within its bounds and is bounded when support is incomplete as with discrete random variables. Our objective is to explore (partial) identification as it relates to the selection problem. To this end, we begin with a foundational building block — missing outcome data.

[1] This chapter draws heavily from Manski, 2007, *Identification for Prediction and Decision*, Harvard University Press, and Manski, 2003, Partial Identification of Probability Distributions, Springer-Verlag..

11.1 Missing outcomes

Suppose each individual in a population is described by the triple (x, y, z) where y is outcome, x is covariates, and $z = 1$ if data are observable and $z = 0$ otherwise. The analyst wishes to describe the unknown distribution for outcome given (random) sample evidence so the analyst poses the question regarding the likelihood y belongs to a subpopulation B. By the law of total probability

$$\begin{aligned}\Pr(y \in B \mid x) &= \Pr(y \in B \mid x, z=1)\Pr(z=1 \mid x) \\ &\quad + \Pr(y \in B \mid x, z=0)\Pr(z=0 \mid x)\end{aligned}$$

The sampling process is informative for $\Pr(y \in B \mid x, z=1)$ and $\Pr(z \mid x)$ but not $\Pr(y \in B \mid x, z=0)$. However, $\Pr(y \in B \mid x, z=0)$ is bounded between 0 and 1. Hence, the identification region is

$$\begin{aligned}& \Pr(y \in B \mid x, z=1)\Pr(z=1 \mid x) \\ \leq \;& \Pr(y \in B \mid x) \leq \\ & \Pr(y \in B \mid x, z=1)\Pr(z=1 \mid x) + \Pr(z=0 \mid x)\end{aligned}$$

The lower bound is the value of $\Pr(y \in B \mid x)$ if missing values never fall in B, and the upper bound is the value of $\Pr(y \in B \mid x)$ if all missing values fall in B. In other words, empirical evidence is informative unless y is always missing.

11.1.1 Examples

Example 1 (identification from the data only) *Suppose we're interested in the likelihood assets are valued between* 90 *and* 110 *(i.e.,* $B \in [90, 110]$*) conditional on covariates* x*. The value of assets is observed when traded* $z = 1$*. However, untraded asset values are unobserved* $z = 0$*. A representative sample of* 190 *assets satisfying the covariate conditions involves* 100 *trades with* 96 *out of* 100 *in the interval* $[90, 110]$*. Hence, taking the sample evidence as representative the identification region is*

$$\begin{aligned}\frac{96}{100}\frac{100}{190} &\leq \Pr(y \in B \mid x) \leq \frac{96}{190} + \frac{90}{190} \\ \frac{96}{190} &\leq \Pr(y \in B \mid x) \leq \frac{186}{190} \\ 0.5053 &\leq \Pr(y \in B \mid x) \leq 0.9789\end{aligned}$$

Here, we've treated the sample evidence as "perfectly" representative of the population — sampling variation (sample variation from the population) is ignored. Even though the interval is wide, it is substantially narrower than the uninformed interval $[0, 1]$.

It may be possible, albeit challenging and perhaps controversial, to tighten the identification region via background knowledge of the market structure. For instance, untraded assets may involve lower bids than the (observed) traded prices, that is, $(y \mid x, z = 1) > (y \mid x, z = 0)$. Alternatively, the retention value to the seller (asking price) may be greater than the value of (observed) traded assets, $(y \mid x, z = 1) < (y \mid x, z = 0)$. If trades of comparable assets are observed,[2] the prevailing condition could be exploited to narrow the identification region. For traded values below the lower bound of B, the $(y \mid x, z = 1) > (y \mid x, z = 0)$ condition makes it less likely missing values are in B. Similarly, traded values above the upper bound of B, the $(y \mid x, z = 1) < (y \mid x, z = 0)$ condition makes it less likely missing values are in B. In other words, background knowledge may indicate $\Pr(y \in B \mid x, z = 0) < 1$ narrowing the identification region. To solidify the point, consider so-called point identification. Stronger background knowledge allowing assignment of the probability distribution may completely define the objective, $\Pr(y \in B \mid x)$.

Example 2 (additional background knowledge) *Continuing with the example 1 setting, suppose the analyst knows y given x is uniformly distributed between 85 and 110. Then, the objective is point-identified.*

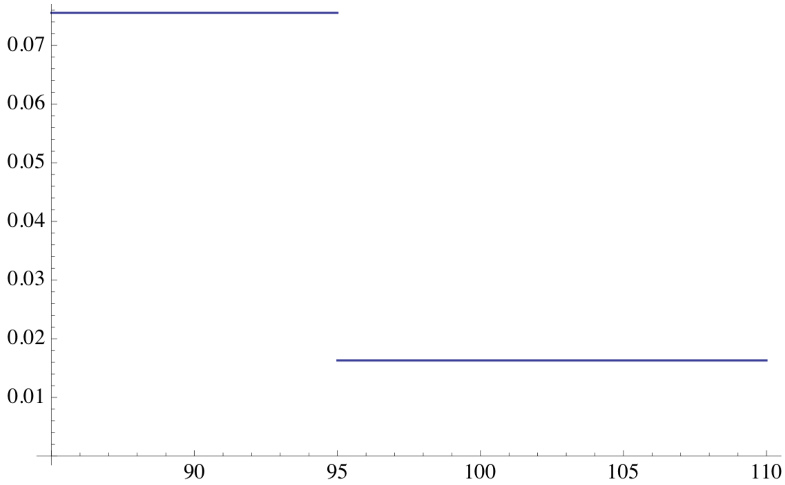
$$\Pr(y \in B \mid x, U) = \Pr(90 \leq y \leq 110 \mid x, U) = \frac{20}{25} = 0.8$$

which, of course, is contained in the region identified by the data alone.

There are numerous consistent conditional distributions. For example,

$$f(y \mid z = 0, x) = \begin{matrix} \frac{17}{225} & 85 \leq y \leq 95 \\ \frac{11}{675} & 95 < y \leq 110 \end{matrix}$$

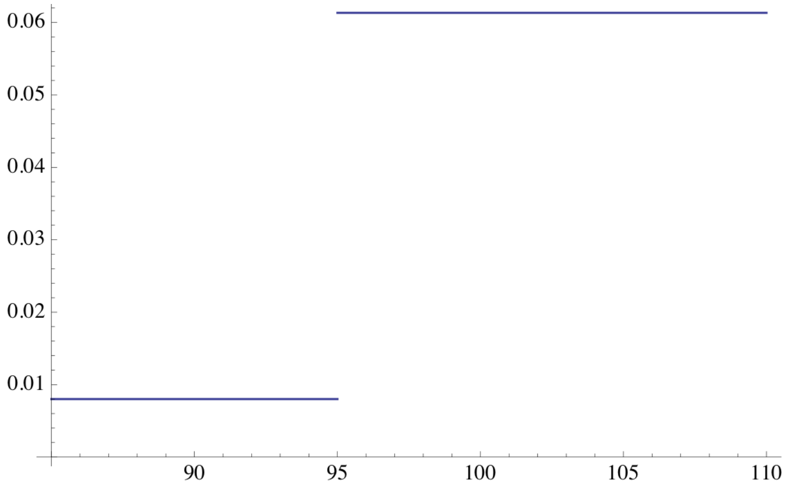
[2] This supposes that assets with the same state-for-state payoffs can be identified such as employed for valuation of redundant securities. However, if the result of this thought experiment is known, we might question the utility of data experiments.



$f\left(y \mid z = 0, x\right)$

and

$$f\left(y \mid z = 1, x\right) = \begin{array}{ll} \frac{1}{125} & 85 \leq y \leq 95 \\ \frac{23}{375} & 95 < y \leq 110 \end{array}$$



$f\left(y \mid z = 1, x\right)$

Then,

$$f\left(y \mid x\right) = \begin{array}{ll} \frac{90}{190}\frac{17}{225} + \frac{100}{190}\frac{1}{125} = \frac{1}{25} & 85 \leq y \leq 95 \\ \frac{90}{190}\frac{11}{675} + \frac{100}{190}\frac{23}{375} = \frac{1}{25} & 95 < y \leq 110 \end{array}$$

with $\Pr(85 \leq y \leq 90 \mid z = 1, x) = 0.04$.

However, if one or both of the uniform bounds are unknown, then the analyst faces an inference problem in which the evidence is informative of the bounds. We return to this example in section 11.1.14 where we discuss Bayesian inference of a probability region based on an assigned uniform likelihood with unknown bounds.

11.1.2 Quantiles

The bounds on the identification region

$$\begin{array}{ll} & \Pr(y \leq t \mid x, z = 1) \Pr(z = 1 \mid x) \\ \leq & \Pr(y \leq t \mid x) \leq \\ & \Pr(y \leq t \mid x, z = 1) \Pr(z = 1 \mid x) + \Pr(z = 0 \mid x) \end{array}$$

translate into α-quantile, $Q_\alpha(y \mid x)$, bounds under some mild regulatory conditions. If the observable probability, $\Pr(y \leq t \mid x, z = 1)$, is continuous and monotone in t, and $\Pr(z = 0 \mid x) < \alpha$ then solving the upper bound for t identifies the least sharp bound for $\Pr(y \mid x, z = 1)$.

$$\Pr(y \leq t \mid x, z = 1) \Pr(z = 1 \mid x) + \Pr(z = 0 \mid x) = \alpha$$

In other words, the α-quantile for $\Pr(y \leq t \mid x)$ is the $\frac{\alpha - \Pr(z=0|x)}{\Pr(z=1|x)}$-quantile of $\Pr(y \mid x, z = 1)$. If $\Pr(z = 0 \mid x) < 1 - \alpha$, solving the lower bound for t identifies the maximum feasible bound for $\Pr(y \mid x, z = 1)$, that is, the $\frac{\alpha}{\Pr(z=1|x)}$−quantile for $\Pr(y \mid x, z = 1)$. Hence, the α-quantile identification region is

$$\begin{array}{ll} & \dfrac{\alpha - \Pr(z = 0 \mid x)}{\Pr(z = 1 \mid x)}\text{-quantile of } \Pr(y \mid x, z = 1) \\ \leq & Q_\alpha(y \mid x) \leq \\ & \dfrac{\alpha}{\Pr(z = 1 \mid x)}\text{-quantile of } \Pr(y \mid x, z = 1) \end{array}$$

More generally, there may be discontinuities and/or flat spots in the observed distribution $\Pr(y \mid x, z = 1)$, Manski [1994] shows that the α-quantile bounds are

$$\begin{array}{ll} & r(\alpha, x) \\ \leq & Q_\alpha(y \mid x) \leq \\ & s(\alpha, x) \end{array}$$

where

$$r(\alpha, x) = \begin{array}{cc} \frac{\alpha - \Pr(z=0|x)}{\Pr(z=1|x)}\text{-quantile of } \Pr(y \mid x, z = 1) & \text{if } \Pr(z = 0 \mid x) < \alpha \\ \min(y) & \text{otherwise} \end{array}$$

and

$$s(\alpha,x) = \begin{array}{cc} \frac{\alpha}{\Pr(z=1|x)}\text{-quantile of } \Pr(y \mid x, z=1) & \text{if } \Pr(z=0 \mid x) < 1-\alpha \\ \max(y) & \text{otherwise} \end{array}$$

11.1.3 Examples

Example 3 (median — data informative) *Suppose we're interested in the median (*$\alpha = 0.5$*) and* $\Pr(z=0 \mid x) = 0.2$*. Then, the* $\alpha = 0.5$*-quantile bounds are* $\frac{0.5-0.2}{0.8}$*-quantile and* $\frac{0.5}{0.8}$*-quantile of* $\Pr(y \mid x, z=1)$*.*

$$\begin{array}{cl} & \frac{3}{8}\textit{-quantile of } \Pr(y \mid x, z=1) \\ \leq & Q_{0.5}(y \mid x) \leq \\ & \frac{5}{8}\textit{-quantile of } \Pr(y \mid x, z=1) \end{array}$$

Further, suppose the observable distribution is uniform $(0,10)$*. Then,*

$$3.75 \leq Q_{0.5}(y \mid x) \leq 6.25$$

The data are informative of the median.

Example 4 (median — data uninformative) *Again, suppose we're interested in the median (*$\alpha = 0.5$*) but* $\Pr(z=0 \mid x) = 0.6$*. Then, the* $\alpha = 0.5$*-quantile bounds are the logical extremes,* $\min(y)$ *and* $\max(y)$*.*

$$\begin{array}{cl} & 0\textit{-quantile of } \Pr(y \mid x) \\ \leq & Q_{0.5}(y \mid x) \leq \\ & 1\textit{-quantile of } \Pr(y \mid x) \end{array}$$

Or, in the case where support for the distribution is $[0,10]$*. Then,*

$$0 \leq Q_{0.5}(y \mid x) \leq 10$$

The data are uninformative of the median.

11.1.4 Expected values

Now, we explore identification from the data alone of the expected value of a function $E[g(y) \mid x]$ in the face of missing data. By the law of iterated expectations,

$$\begin{array}{rcl} E[g(y) \mid x] & = & E[g(y) \mid x, z=1]\Pr(z=1 \mid x) \\ & & +E[g(y) \mid x, z=0]\Pr(z=0 \mid x) \end{array}$$

From the data alone, the identification region for the mean is

$$\begin{aligned} & E\left[g\left(y\right)\mid x,z=1\right]\Pr\left(z=1\mid x\right)+g_{0}\Pr\left(z=0\mid x\right) \\ \leq\ & E\left[g\left(y\right)\mid x\right]\leq \\ & E\left[g\left(y\right)\mid x,z=1\right]\Pr\left(z=1\mid x\right)+g_{1}\Pr\left(z=0\mid x\right) \end{aligned}$$

where $g_0 = \min g\left(y\right)$ and $g_1 = \max g\left(y\right)$. Whenever $\Pr\left(z=0\mid x\right)<1$ and $g\left(y\right)$ is bounded, the data are informative. The width of the identification region is $\left(g_1-g_0\right)\Pr\left(z=0\mid x\right)$. However, if either extreme is unbounded, $g_0=-\infty$ or $g_1=\infty$, the identification has region has infinite width. The data remain informative for the mean so long as $g\left(y\right)$ has at least one finite bound.

Next, we explore some examples. First, we illustrate some standard point-identification strategies for addressing means with missing outcome data. Then, we return to partial identification strategies for means with missing outcome data.

11.1.5 Point identification of means

We briefly illustrate two missing at random (MAR) strategies: inverse probability weighting (IPW) and imputation. And, a non-MAR strategy: Heckman's control function.

11.1.6 Examples

Example 5 ($\boldsymbol{IPW}$) *Suppose the DGP is*

y	x	z
3	2	1
5	2	1
4	2	0
1	1	1
0	1	0
2	1	0

From the DGP, $E\left[y\mid x=2\right]=4$*,* $E\left[y\mid x=1\right]=1$*,* $E\left[y\right]=\frac{5}{2}$*,* $n_{x=2}=3$*,* $n_{x=1}=3$*,* $\Pr\left(z=1\mid x=2\right)=E\left[z\mid x=2\right]=\frac{2}{3}$*, and* $E\left[z\mid x=1\right]=\frac{1}{3}$*.*[3] *As the name suggests, IPW utilizes the observed data scaled by the inverse of the propensity score to identify the unknown mean for outcome.*

$$E\left[y\mid x\right]=\frac{1}{n_x}\sum\frac{y\cdot 1\left(z=1\mid x\right)}{\Pr\left(z=1\mid x\right)}$$

[3] To conserve space, we refrain from writing all permutations for the DGP. For completeness, the DGP has nine values for each level of x with outcomes equally likely and all outcome values are equally likely to be missing given x.

where $1(\cdot)$ *is an indicator function equal to one when the condition inside is satisfied and zero otherwise. In other words,*

$$E[y \mid x=2] = \frac{1}{3}\left(\frac{3}{\frac{2}{3}} + \frac{5}{\frac{2}{3}}\right) = 4$$

and

$$E[y \mid x=1] = \frac{1}{3}\left(\frac{1}{\frac{1}{3}}\right) = 1$$

By iterated expectations, the unconditional outcome mean is

$$\begin{aligned} E[y] &= \Pr(x=2)\, E[y \mid x=2] + \Pr(x=1)\, E[y \mid x=1] \\ &= \frac{1}{2}(4) + \frac{1}{2}(1) = \frac{5}{2} \end{aligned}$$

where $\Pr(x=i) = \frac{n_i}{n_2+n_1}$, $i = 1, 2$.

Example 6 (imputation) *Suppose the DGP is the same as example 5. Some perhaps flexible model is employed to fit the conditional means based on the observed data yielding*

$$\begin{aligned} m(x=2) &= 4 \\ m(x=1) &= 1 \end{aligned}$$

Imputation identifies conditional outcome means as

$$E[y \mid x] = \frac{1}{n_x} \sum \{ y \cdot 1(z=1 \mid x) + m(x) \cdot 1(z=0 \mid x) \}$$

That is,

$$E[y \mid x=2] = \frac{1}{3}\{3+5+(4)\} = 4$$

and

$$E[y \mid x=1] = \frac{1}{3}\{1+(1+1)\} = 1$$

where imputed missing values are in parentheses. Unconditional outcome expectations are again

$$\begin{aligned} E[y] &= \Pr(x=2)\, E[y \mid x=2] + \Pr(x=1)\, E[y \mid x=1] \\ &= \frac{1}{2}(4) + \frac{1}{2}(1) = \frac{5}{2} \end{aligned}$$

When data are missing by choice rather than missing at random the identification challenge is considerably more daunting. Heckman's strategy employs strong identifying conditions: namely, normality of unobservable utility underlying the choice model and exclusion restrictions for the instruments w. Then, by properties of the truncated normal the analyst extrapolates into the truncated tail region to identify expected values for both the observed (by choice) and the missing outcome data. The mean (without covariates)[4] for the observable data is identified by

$$E\left[y \mid z=1\right] = \mu + \rho\sigma\lambda\left(z=1\right)$$

while the mean for missing data is identified by

$$E\left[y \mid z=0\right] = \mu + \rho\sigma\lambda\left(z=0\right)$$

where $\lambda\left(z=1\right) = \frac{\phi(w\theta)}{\Phi(w\theta)}$ and $\lambda\left(z=0\right) = -\frac{\phi(w\theta)}{1-\Phi(w\theta)}$. The estimation strategy is in two-steps. First, regress z onto w via probit. Then, regress observed outcome $\left(y \mid z=1\right)$ onto an intercept and estimated $\lambda\left(z=1\right)$ based on results from the first stage probit. Then, plug-in the estimates for μ, $\rho\sigma$, and $\lambda\left(z\right)$ in the above expressions to get the estimated values for $E\left[y \mid z=1\right]$ and $E\left[y \mid z=0\right]$. The example below is discrete rather than continuous but (hopefully) compactly illustrates the ideas.

Example 7 (Heckman's control function strategy) *Suppose the DGP is*

y	$y-\mu$	z	w	$\rho\sigma\lambda\left(z\right)$
15	3	1	−2.5	3
14	2	1	1	2
13	1	1	5.5	1
11	−1	0	−5.5	−1
10	−2	0	−1	−2
9	−3	0	2.5	−3

where $E\left[w^T\left(y-\mu\right)\right] = 0$. *The expected value for the observed (by choice) outcomes is*

$$E\left[y \mid z=1\right] = \frac{1}{3}\left\{(12+3)+(12+2)+(12+1)\right\} = 14$$

while the expected value for missing outcomes is

$$E\left[y \mid z=0\right] = \frac{1}{3}\left\{(12-1)+(12-2)+(12-3)\right\} = 10$$

and by iterated expectations the unconditional expectation is

$$\begin{aligned} E\left[y\right] &= \Pr\left(z=1\right)E\left[y \mid z=1\right] + \Pr\left(z=0\right)E\left[y \mid z=0\right] \\ &= \frac{1}{2}\left(14\right) + \frac{1}{2}\left(10\right) = 12 \end{aligned}$$

[4] If there are covariates, then simply include their relation with outcome, say $E\left[y \mid x,z\right] = \mu + g\left(x\right) + \rho\sigma\lambda\left(z\right)$.

11.1.7 Partial identification of means

Next, we illustrate partial identification of expected values of functions of random variables. First, we consider the expected value of a random variable with bounded support. Then, we illustrate how the data are informative for the variance.

11.1.8 Examples

Example 8 (mean) *Suppose we're interested in the mean of y where the evidence indicates $E[y \mid x, z = 1] = 3$, $\Pr(z = 0 \mid x) = 0.4$, and known support for y is $[-10, 10]$. Then, the identification region for the mean is*

$$\begin{aligned} 3(0.6) - 10(0.4) &\leq E[y \mid x] \leq 3(0.6) + 10(0.4) \\ -2.2 &\leq E[y \mid x] \leq 5.8 \end{aligned}$$

Example 9 (variance) *Suppose we're interested in the variance of y where the evidence indicates $E[g(y) \mid x, z = 1] = E\left[(y - E[y])^2 \mid x, z = 1\right] = 100$, $\Pr(z = 0 \mid x) = 0.4$, and support for $g(y)$ is unbounded above $(0, \infty)$. Then, the identification region for the variance is*

$$\begin{aligned} 100(0.6) + 0(0.4) &\leq Var[y \mid x] \leq 100(0.6) + \infty(0.4) \\ 60 &\leq Var[y \mid x] < \infty \end{aligned}$$

The data are informative as the interval is narrower than the maximum, $0 \leq Var[y \mid x] < \infty$.

11.1.9 Inference

We briefly discuss inference for a partially-identified unknown mean, θ, from a random sample of size n. Suppose outcome is known to have $[0, 1]$ support where $E[y \mid z = 1] = \mu_1 \in [0, 1]$, $E[y \mid z = 0] = \mu_0 \in [0, 1]$, $Var[y \mid z = 1] = \sigma^2$, $n_1 = \sum_{i=1}^{n} z_i$, $\overline{y} = \frac{\sum_i y_i z_i}{\sum_i z_i}$, and, to keep matters simple, propensity score[5] $E[z \mid x] = p$ is known.[6] Bounds on the mean are

$$pE[y \mid z = 1] \leq E[y] = \theta \leq pE[y \mid z = 1] + (1 - p)$$

Replacing parameters with sample analogs gives

$$p\overline{y} \leq \theta \leq p\overline{y} + (1 - p)$$

[5] Conditioning on covariates, x, is made explicit to emphasize the propensity score but otherwise suppressed to simplify notation.

[6] Imbens and Manski [2004] show, in the general case when σ and p are unknown, σ^2 can be replaced by $\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} z_i (y_i - \overline{y})^2}{n_1 - 1}$ and p can be replaced by $\widehat{p} = \frac{\sum_{i=1}^{n} z_i}{n}$.

Asymptotic normality of the sample average produces a 95% confidence interval for the lower bound

$$\left[p\left(\overline{y}-1.96\frac{\sigma}{\sqrt{n_1}}\right),p\left(\overline{y}+1.96\frac{\sigma}{\sqrt{n_1}}\right)\right]$$

and for the upper bound

$$\left[p\left(\overline{y}-1.96\frac{\sigma}{\sqrt{n_1}}\right)+(1-p),p\left(\overline{y}+1.96\frac{\sigma}{\sqrt{n_1}}\right)+(1-p)\right]$$

A valid, conservative confidence interval for the mean utilizes the lower confidence limit from the lower bound and upper confidence limit from the upper bound.

$$\left[p\left(\overline{y}-1.96\frac{\sigma}{\sqrt{n_1}}\right),p\left(\overline{y}+1.96\frac{\sigma}{\sqrt{n_1}}\right)+(1-p)\right]$$

Imbens and Manski (*Econometrica* [2004]) suggest a confidence interval that supplies uniform, say $\alpha = 95\%$, coverage for all p.

$$\left[p\left(\overline{y}-C_n\frac{\sigma}{\sqrt{n_1}}\right),p\left(\overline{y}+C_n\frac{\sigma}{\sqrt{n_1}}\right)+(1-p)\right]$$

where C_n solves

$$\alpha=\Phi\left(C_n+\sqrt{n\widehat{p}}\frac{1-p}{\sigma p}\right)-\Phi\left(-C_n\right)$$

for $\Phi(\cdot)$ the standard normal CDF and $\widehat{p}=\frac{\sum_{i=1}^{n}z_i}{n}$. For $p=1$ (the mean is point-identified) and $\alpha=.95$, $C_n=1.96$ and as indicated by the examples below for $p<1$, C_n decreases towards 1.645.

To see this, recognize $(\overline{y},\widehat{p})$ are a pair of sufficient statistics for θ and $(\overline{y}\mid\widehat{p})\sim N\left(\mu_1,\frac{\sigma^2}{\widehat{p}n}\right)$ or, at least, $\sqrt{n}\left(\overline{y}-\mu_1\mid\widehat{p}\right)\xrightarrow{d}N\left(0,\frac{\sigma^2}{\widehat{p}}\right)$.[7] Symmetric intervals for partially-identified θ are of the form

$$\left[\widehat{\theta}_l-D,\widehat{\theta}_u+D\right]$$

or

$$\Pr\left(\widehat{\theta}_l-D\leq\theta\leq\widehat{\theta}_u+D\mid\widehat{p}\right)$$

Substituting upper and lower bounds gives

$$\Pr\left(p\overline{y}-D\leq p\mu_1+(1-p)\,\mu_0\leq p\overline{y}+(1-p)+D\mid\widehat{p}\right)$$

[7] If the observed outcome distribution is unknown or unassigned, the sample average converges asymptotically to a normal distribution via the central limit theorem.

Collecting terms involving observable data produces

$$\Pr\left(-D-(1-p)\,\mu_0 \le p\,(\mu_1-\overline{y}) \le (1-\mu_0)\,(1-p)+D \mid \widehat{p}\right)$$

Now, rescale by the standard deviation of $(\overline{y} \mid \widehat{p})$ and divide by p.

$$\Pr\left(\frac{-D-(1-p)\,\mu_0}{p\frac{\sigma}{\sqrt{\widehat{p}n}}} \le \frac{p\,(\mu_1-\overline{y})}{p\frac{\sigma}{\sqrt{\widehat{p}n}}} \le \frac{(1-\mu_0)\,(1-p)+D}{p\frac{\sigma}{\sqrt{\widehat{p}n}}} \mid \widehat{p}\right)$$

The probability is the same at either endpoint $\mu_0 = 0, 1$ and a global minimum at the extremes as the second derivative is negative for all $\mu_0 \in [0,1]$. Let $C_n = D\frac{\sqrt{n\widehat{p}}}{p\sigma}$ and set the probability coverage equal to α, then we have

$$\alpha = \Pr\left(-C_n \le \frac{(\mu_1-\overline{y})}{\frac{\sigma}{\sqrt{\widehat{p}n}}} \le C_n + \sqrt{n\widehat{p}}\frac{1-p}{\sigma p} \mid \widehat{p}\right)$$

as claimed above. Both the conservative interval and Imbens-Manski uniform interval converge to the identification region asymptotically.

11.1.10 Examples

Example 10 (confidence intervals for a partially-identified mean) *Suppose the DGP and sample size for various instances of missing outcomes are*

Pr	y	z_1	z_2	z_3	z_4
0.035	0	1	1	1	1
0.05	$\frac{1}{9}$	1	1	0	1
0.09	$\frac{2}{9}$	1	1	1	0
0.15	$\frac{3}{9}$	1	1	1	1
0.175	$\frac{4}{9}$	1	1	1	0
0.175	$\frac{5}{9}$	1	1	1	1
0.15	$\frac{6}{9}$	1	1	1	1
0.09	$\frac{7}{9}$	1	1	1	0
0.05	$\frac{8}{9}$	1	1	0	1
0.035	1	1	0	1	1
μ_1	0.5	0.5	0.482	0.5	0.515
σ	0.2365	0.2365	0.2204	0.2129	0.2532
p		1	0.9	0.8	0.7
n_1		100	90	80	70

Suppose the sample information is representative of the population ($\overline{y} = \mu_1$ and $\widehat{p} = p$). The conservative and Imbens-Manski 95% confidence intervals *are*

	conservative 95% *confidence intervals*	
case	*lower limit*	*upper limit*
z_1	$1\left(0.5-1.96\frac{0.2365}{\sqrt{100}}\right)$ $=0.454$	$1\left(0.5+1.96\frac{0.2365}{\sqrt{100}}\right)+0$ $=0.546$
z_2	$0.9\left(0.482-1.96\frac{0.2204}{\sqrt{90}}\right)$ $=0.393$	$0.9\left(0.482+1.96\frac{0.2204}{\sqrt{90}}\right)+0.1$ $=0.575$
z_3	$0.8\left(0.5-1.96\frac{0.2129}{\sqrt{80}}\right)$ $=0.363$	$0.8\left(0.5+1.96\frac{0.2129}{\sqrt{80}}\right)+0.2$ $=0.637$
z_4	$0.7\left(0.515-1.96\frac{0.2532}{\sqrt{70}}\right)$ $=0.319$	$0.7\left(0.515+1.96\frac{0.2532}{\sqrt{70}}\right)+0.3$ $=0.702$

	Imbens-Manski 95% *confidence intervals*	
case	*lower limit*	*upper limit*
z_1	$1\left(0.5-1.96\frac{0.2365}{\sqrt{100}}\right)$ $=0.454$	$1\left(0.5+1.96\frac{0.2365}{\sqrt{100}}\right)+0$ $=0.546$
z_2	$0.9\left(0.482-1.645\frac{0.2204}{\sqrt{90}}\right)$ $=0.399$	$0.9\left(0.482+1.645\frac{0.2204}{\sqrt{90}}\right)+0.1$ $=0.568$
z_3	$0.8\left(0.5-1.645\frac{0.2129}{\sqrt{80}}\right)$ $=0.369$	$0.8\left(0.5+1.645\frac{0.2129}{\sqrt{80}}\right)+0.2$ $=0.631$
z_4	$0.7\left(0.515-1.645\frac{0.2532}{\sqrt{70}}\right)$ $=0.326$	$0.7\left(0.515+1.645\frac{0.2532}{\sqrt{70}}\right)+0.3$ $=0.695$

Example 11 (Imbens-Manski confidence intervals) *Suppose the DGP is one of the four cases below and again the sample information is representative of the population (*$\overline{y}=\mu_1$ *and* $\widehat{p}=p$*)..*

	case 1	*case 2*	*case 3*	*case 4*
μ_1	0.5	0.5	0.482	0.482
σ	0.2365	0.2365	0.2204	0.2204
p	0.99	0.95	0.99	0.95
n_1	99	95	99	95

The following table depicts the variation in the critical value, C_n, to illustrate dependence of the critical value on σ as well as p.

	Imbens-Manski 95% *confidence intervals*	
case	*lower limit*	*upper limit*
1	$0.99\left(0.5 - 1.791\frac{0.2365}{\sqrt{99}}\right)$ $= 0.453$	$0.99\left(0.5 + 1.791\frac{0.2365}{\sqrt{99}}\right) + 0.01$ $= 0.547$
2	$0.95\left(0.482 - 1.6455\frac{0.2365}{\sqrt{95}}\right)$ $= 0.434$	$0.95\left(0.482 + 1.6455\frac{0.2365}{\sqrt{95}}\right) + 0.05$ $= 0.566$
3	$0.99\left(0.482 - 1.782\frac{0.2204}{\sqrt{99}}\right)$ $= 0.438$	$0.99\left(0.482 + 1.782\frac{0.2204}{\sqrt{99}}\right) + 0.01$ $= 0.526$
4	$0.95\left(0.482 - 1.6452\frac{0.2204}{\sqrt{95}}\right)$ $= 0.419$	$0.95\left(0.482 + 1.6452\frac{0.2204}{\sqrt{95}}\right) + 0.05$ $= 0.546$

11.1.11 Respecting stochastic dominance

Parameters that respect stochastic dominance often enable relaxed or more credible identifying conditions such as some form of monotonicity (discussed later). Distribution Q stochastically dominates distribution Q' if $Q\left(y \leq t\right) \leq Q^{'}\left(y \leq t\right)$ for all t. A parameter $D\left(\cdot\right)$ is said to *respect stochastic dominance* if $D\left(Q\right) \geq D\left(Q^{'}\right)$ whenever Q stochastically dominates Q'. Sharp bounds are straightforwardly defined for parameters that respect stochastic dominance. For lower bounds, missing values are replaced by g_0, and for upper bounds, missing values are replaced by g_1. Quantiles and means of increasing functions of y respect stochastic dominance but spread parameters such as variance and interquartile range do not (see Blundell et al (BGIM) [2007]).[8]

11.1.12 Examples

Example 12 (mean respects stochastic dominance) *Suppose the data are distributed as in the table below where $Q\left(y\right)$ stochastically dominates $Q^{'}\left(y\right)$.*

y	$Q^{'}\left(y\right)$	$Q\left(y\right)$
0	$\frac{2}{6}$	0
1	$\frac{3}{6}$	$\frac{3}{6}$
2	$\frac{5}{6}$	$\frac{4}{6}$
3	1	1

[8] See the appendix for a brief discussion of BGIM's bounds for spread parameters.

The mean respects stochastic dominance: $E_Q[y] = \frac{11}{6} > E_{Q'}[y] = \frac{8}{6}$*.*

Example 13 (variance doesn't respect stochastic dominance) *The data continue to be distributed as in example 12. Variance does not respect stochastic dominance:* $Var_Q[y] = \frac{29}{36} < Var_{Q'}[y] = \frac{11}{9}$*.* Q *stochastically dominates* Q'*, the mean for* Q *is larger than the mean for* Q'*, but the variance of* Q *is smaller than the variance for* Q'*.*

11.1.13 Distributional assumptions

Identification of unbounded random variables requires distributional assumptions. Though some analysts shy away from the idea, maximum entropy probability assignment can be an illuminating strategy. Maximum entropy (Jaynes [2004]) suggests that the analyst assign probabilities based on background knowledge of the setting but no more — hence, maximum entropy (uncertainty or disorder) subject to what one knows (typically expressed in terms of moment conditions; see chapter 4).

Briefly, probabilities are assigned by solving the Lagrangian for entropy.

$$\begin{array}{rl}\max\limits_{p_i \geq 0} & -\sum_{i=1}^{n} p_i \log p_i - \lambda_0 \left(\sum_{i=1}^{n} p_i - 1\right) \\ & -\lambda_1 \left(\text{moment condition}_1\right) - \ldots \\ & -\lambda_n \left(\text{moment condition}_n\right)\end{array}$$

This problem can be framed as assigning a kernel, k_i, and then normalizing the kernel to identify a proper distribution.[9]

$$k_i = \exp\left[-\lambda_1 \frac{\partial \left(\text{moment condition}_1\right)}{\partial p_i} - \ldots - \lambda_n \frac{\partial \left(\text{moment condition}_n\right)}{\partial p_i}\right]$$

Let Z be a normalizing or partition function[10]

$$Z = \sum_{i=1}^{n} k_i$$

then probabilities are assigned as

$$p_i = \tfrac{k_i}{Z}, \quad \text{for all } i = 1, \ldots, n$$

11.1.14 Probability assignment and Bayesian inference

Now, we revisit example 1 involving missing asset value outcomes and discuss Bayesian inference when distribution bounds are unknown.

[9] For continuous rather than discrete support, a probability mass, p_i, is replaced with a density function, $f(x_i)$, and summation is replaced with integration which effectively replaces entropy with differential entropy.

[10] Notice λ_0 and its associated constraint is absorbed in the partition function.

Example 14 (uniform with unknown upper bound) *Suppose the missing data in example 1 are believed to have the same support as the observed and the market structure indicates asset value is bounded below at* 85 *but the upper bound is unknown. Also, the analyst deems it credible to assign a uniform distribution with unknown upper bound, u, and lower bound* 85 *for outcome.*[11] *In other words, the analyst is only confident of the lower bound. Further, suppose the analyst believes the distribution for the unknown upper bound is no smaller than* 100, *and the distribution is relatively flat such that the expected value of its natural logarithm is very large, say* 1002.71. *Then, the maximum entropy prior for the upper bound u is conjugate prior to uniform data, a Pareto distribution with hyperparameters* $a = 0.001$ *and* $b = 100$.

$$p(u) = \frac{a(b-85)^{a}}{(u-85)^{a+1}} = \frac{0.00100271}{(u-85)^{1.001}}$$

The likelihood is

$$\ell(u \mid y_1, x) \propto \frac{1}{(u-85)^{n}} = \frac{1}{(u-85)^{100}}$$

where y_1 *refers to observed outcomes. Hence, the posterior distribution for the upper bound is also a Pareto distribution with hyperparameters* $a + n = 100.001$ *and* $\max[b, y^u] = 108$.

$$p(u \mid y_1, x) = \tfrac{(a+n)(\max[b,y^u]-85)^{a+n}}{(u-85)^{a+n+1}}, \quad \max[b, y^u] \leq u \leq \infty$$

or

$$p(u \mid y_1, x) = \tfrac{1.49331*10^{138}}{(u-85)^{101.001}}, \quad 108 \leq u \leq \infty$$

where $y^u = 108$ *is the maximum y in the sample. The posterior distribution for u is concentrated between* 108 *and* 110

$$\Pr(108 \leq u \leq 110 \mid y_1, x) = 0.999761$$

and

$$E[u \mid y_1, x] = 108.232$$

Then, on average, $\Pr(y \in B \mid y_1, x)$ *equals*

$$\begin{aligned} E[\Pr(90 \leq y \leq 110 \mid y_1, x)] &\equiv E\left[\frac{\min(110, u) - 90}{u - 85} \mid y_1, x\right] \\ &= 0.784759 \end{aligned}$$

[11] As discussed earlier, this does not imply outcome conditional on missing or observable have the same uniform distribution. For example, each conditional distribution could be a mixture of distributions.

If the analyst regards $u > 115$ *as an inconsequential likelihood then inference based on this probability assignment in conjunction with the data is almost surely*

$$\frac{2}{3} \leq \Pr\left(y \in B \mid y_1, x\right) \leq \frac{4}{5}$$

These bounds are quite conservative when compared with 1000 *posterior simulation draws of the hyperparameter* u *and* $\Pr\left(90 \leq y \leq 110 \mid y_1, x\right) = \frac{\min(110,u)-90}{u-85}$.

statistic	u	$\Pr\left(90 \leq y \leq 110 \mid y_1, x\right)$
mean	108.227	0.7847
standard deviation	0.239	0.0022
minimum	108.000	0.7826
0.01-*quantile*	108.002	0.7826
0.05-*quantile*	108.010	0.7827
0.10-*quantile*	108.020	0.7828
0.25-*quantile*	108.062	0.7832
0.50-*quantile*	108.145	0.7840
0.75-*quantile*	108.315	0.7855
0.90-*quantile*	108.530	0.7875
0.95-*quantile*	108.684	0.7889
0.99-*quantile*	109.117	0.7847
maximum	109.638	0.7971

Bayesian inference with unknown upper bound

This inference region afforded by an assigned likelihood in combination with an assigned prior over the unknown upper bound addresses sampling variation and is considerably narrower than the identification region based on the data alone where we've ignored sampling variation. This affirms the considerable identifying power of probability distribution assignment.

Example 15 (uniform with both bounds unknown) *Again, suppose the missing data are believed to have the same support as the observed and both the lower and upper bounds are unknown. Also, the analyst deems it credible only to assign a uniform distribution with unknown upper bound,* u*, and lower bound,* l *for outcome. Further, suppose the analyst believes the distribution for the unknown lower bound is no larger than* 99*, for the unknown upper bound is no smaller than* 100*, and the distribution is relatively flat such that the expected value of its natural logarithm is very large,* 1001*. Then, the maximum entropy prior for the bounds* l *and* u *is conjugate prior to uniform data, a bilateral, bivariate Pareto distribution with hyperparameters* $a = 0.001$*,* $r_1 = 99$*, and* $r_2 = 100$*.*

$$p\left(u, l\right) = \frac{a\left(a+1\right)\left(r_2 - r_1\right)^a}{\left(u-l\right)^{a+2}} = \frac{0.001001}{\left(u-l\right)^{2.001}}$$

The likelihood is

$$\ell\left(l.u \mid y_1, x\right) \propto \frac{1}{\left(u-l\right)^n} = \frac{1}{\left(u-l\right)^{100}}$$

Hence, the posterior distribution for the upper bound is also a bivariate Pareto distribution with hyperparameters $\min\left[r_1, y^l\right] = 85$*,* $\max\left[r_2, y^u\right] = 108$*, and* $a+n = 100.001$*.*

$$\begin{aligned} &p\left(l, u \mid y_1, x\right) \\ &\quad = \frac{(a+n)(a+n+1)\left(\max[r_2,y^u]-\min\left[r_1,y^l\right]\right)^{a+n}}{(u-l)^{a+n+2}}, \qquad \begin{array}{c} -\infty \le l \le \min\left[r_1, y^l\right], \\ \max\left[r_2, y^u\right] \le u \le \infty \end{array} \end{aligned}$$

or

$$p\left(l, u \mid y_1, x\right) = \frac{1.50826 * 10^{140}}{(u-l)^{102.001}}, \qquad \begin{array}{c} -\infty \le l \le 85, \\ 108 \le u \le \infty \end{array}$$

where $y^l = 85$ *is the minimum* y *and* $y^u = 108$ *is the maximum* y *in the sample. The posterior distribution for* l *is concentrated between* 83 *and* 85 *and for* u *is concentrated between* 108 *and* 110*.*

$$\Pr\left(83 \le l \le 85, 108 \le u \le 110 \mid y_1, x\right) = 0.999522$$

and

$$E\left[u - l \mid y_1, x\right] = 108.2323 - 84.7677 = 23.4646$$

such that there is negligible probability that l *is below* 80 *and* u *exceeds* 115*. The expected* $\Pr\left(y \in B \mid x\right)$ *equals* $E\left[\frac{\min(110,u)-\max(90,l)}{u-l} \mid y_1, x\right] = 0.777065$*. If the analyst regards* $l < 80$ *and* $u > 115$ *as an inconsequential likelihood then inference based on this probability assignment in conjunction with the data is almost surely*

$$\frac{4}{7} \le \Pr\left(y \in B \mid y_1, x\right) \le \frac{4}{5}$$

Again, this interval is conservative compared with 1000 *posterior simulation draws.*

statistic	u	l	$\Pr(90 \leq y \leq 110 \mid y_1, x)$
mean	108.344	84.885	0.7819
standard deviation	0.251	0.118	0.0058
minimum	108.005	84.095	0.7530
0.01-*quantile*	108.025	84.474	0.7653
0.05-*quantile*	108.058	84.653	0.7715
0.10-*quantile*	108.088	84.734	0.7745
0.25-*quantile*	108.167	84.847	0.7790
0.50-*quantile*	108.289	84.922	0.7827
0.75-*quantile*	108.457	84.965	0.7857
0.90-*quantile*	108.659	84.985	0.7882
0.95-*quantile*	108.822	84.993	0.7899
0.99-*quantile*	109.267	84.999	0.7939
maximum	109.683	84.9998	0.7974

Bayesian inference with unknown upper and lower bounds

As expected, this inference region is wider than the previous case when the lower bound is known. However, this inference region afforded by an assigned likelihood in combination with an assigned prior over the unknown bounds of support is considerably narrower than the identification region based on the data alone which ignores sampling variation.

11.1.15 Refutability

An important tool sometimes at our disposal is refutability of identifying conditions. Refutability logically derives from the evidence while credibility is a matter of judgment. Any conditions restricting the distribution of missing data are nonrefutable — the data cannot test their veracity. On the other hand, some distributional restrictions are refutable. For example, if the identification region based on the data alone is inconsistent with the restriction on the distribution (i.e., their intersection is empty), then the distributional restriction is refuted.

11.1.16 Example

Example 16 (mean example revisited) *Suppose the analyst assigns a conditional outcome distribution that yields a mean in the region* $[-7, -3]$*. Recall the identification region for the mean based on the data in example 8 is*

$$-2.2 \leq E[y \mid 2] \leq 5.8$$

As the intervals do not intersect, the condition on the mean of the distribution is refuted by the data.

The preceding discussion treats the data as "perfectly" representative of the "observed" population. Of course, typically the analyst employs sample evidence (in the form of the sample analog) to infer the identification region based on the data. Refutable tests then rest on this sample-based identification region lying sufficiently far from the region defined by the distributional condition.

11.1.17 Missing covariates and outcomes

The above discussion focuses on missing outcomes, sometimes both outcomes and covariates are missing. Now, we consider the case where outcomes and covariates are jointly missing. We denote this as $z = 1$ when outcomes and covariates and jointly observed and $z = 0$ when both are missing. Again, by the law of total probability

$$\begin{aligned}\Pr(y \mid x = x_0) &= \Pr(y \mid x = x_0, z = 1) \Pr(z = 1 \mid x = x_0) \\ &\quad + \Pr(y \mid x = x_0, z = 0) \Pr(z = 0 \mid x = x_0)\end{aligned}$$

As before, the sampling process reveals $\Pr(y \mid x = x_0, z = 1)$ but nothing about $\Pr(y \mid x = x_0, z = 0)$. However, in this case the sampling process only partially reveals $\Pr(z \mid x = x_0)$.

From Bayes theorem, we have for $i = 0$ or 1

$$\begin{aligned}&\Pr(z = i \mid x = x_0) \\ &= \frac{\Pr(x = x_0 \mid z = i) \Pr(z = i)}{\Pr(x = x_0 \mid z = 1) \Pr(z = 1) + \Pr(x = x_0 \mid z = 0) \Pr(z = 0)}\end{aligned}$$

Substituting this into the expression for $\Pr(y \mid x = x_0)$ yields

$$\begin{aligned}&\Pr(y \mid x = x_0) \\ &= \frac{\Pr(y \mid x = x_0, z = 1) \Pr(x = x_0 \mid z = 1) \Pr(z = 1)}{\Pr(x = x_0 \mid z = 1) \Pr(z = 1) + \Pr(x = x_0 \mid z = 0) \Pr(z = 0)} \\ &\quad + \frac{\Pr(y \mid x = x_0, z = 0) \Pr(x = x_0 \mid z = 0) \Pr(z = 0)}{\Pr(x = x_0 \mid z = 1) \Pr(z = 1) + \Pr(x = x_0 \mid z = 0) \Pr(z = 0)}\end{aligned}$$

The sampling process reveals $\Pr(x = x_0 \mid z = 1)$, $\Pr(y \mid x = x_0, z = 1)$, and $\Pr(z)$ but offers no evidence on $\Pr(x = x_0 \mid z = 0)$ or $\Pr(y \mid x = x_0, z = 0)$.

$$\begin{aligned}\Pr(y \mid x = x_0) &= \frac{\Pr(y \mid x = x_0, z = 1) \Pr(x = x_0 \mid z = 1) \Pr(z = 1)}{\Pr(x = x_0 \mid z = 1) \Pr(z = 1) + p \Pr(z = 0)} \\ &\quad + \frac{\Pr(y \mid x = x_0, z = 0)\, p \Pr(z = 0)}{\Pr(x = x_0 \mid z = 1) \Pr(z = 1) + p \Pr(z = 0)}\end{aligned}$$

Recognizing $p \equiv \Pr(x = x_0 \mid z = 0) \in [0, 1]$ and the region is largest for $p = 1$ leads to an identification region based on the data alone equal to

$$\begin{aligned} & \frac{\Pr(y \mid x = x_0, z = 1)\Pr(x = x_0 \mid z = 1)\Pr(z = 1)}{\Pr(x = x_0 \mid z = 1)\Pr(z = 1) + \Pr(z = 0)} \\ \leq\ & \Pr(y \mid x = x_0) \leq \\ & \frac{\Pr(y \mid x = x_0, z = 1)\Pr(x = x_0 \mid z = 1)\Pr(z = 1)}{\Pr(x = x_0 \mid z = 1)\Pr(z = 1) + \Pr(z = 0)} \\ & + \frac{\Pr(z = 0)}{\Pr(x = x_0 \mid z = 1)\Pr(z = 1) + \Pr(z = 0)} \\ =\ & \Pr(y \mid x = x_0, z = 1)\gamma + (1 - \gamma) \end{aligned}$$

where $\gamma = \frac{\Pr(x=x_0|z=1)\Pr(z=1)}{\Pr(x=x_0|z=1)\Pr(z=1)+\Pr(z=0)}$. The data are informative so long as $\Pr(y \mid x = x_0, z = 1) > 0$. When $\Pr(y \mid x = x_0, z = 1) = 0$ the lower bound is zero; hence, the data are uninformative. This suggests the identification region based on the data is narrower when only the outcomes are missing than when outcomes and covariates are missing.

11.1.18 Examples

Example 17 (jointly missing outcomes and covariates) *Return to example 1 and compare the cases involving outcomes missing with outcomes and covariates jointly missing. For simplicity, suppose the sampling process satisfies* $\Pr(z){=}\Pr(x = x_0 \mid z = 1) = \frac{100}{190}$, $\Pr(y \in B \mid x = x_0, z = 1) = \frac{96}{100}$, *and* $x = x_0$. *The identification region based on the data when outcomes and covariates are jointly missing is*

$$\begin{aligned} & \frac{96}{100}\frac{\frac{100}{190}\frac{100}{190}}{\frac{100}{190}\frac{100}{190}+\frac{90}{190}} \\ \leq\ & \Pr(y \in B \mid x = x_0) \leq \\ & \frac{96}{100}\frac{\frac{100}{190}\frac{100}{190}}{\frac{100}{190}\frac{100}{190}+\frac{90}{190}} + \frac{\frac{90}{190}}{\frac{100}{190}\frac{100}{190}+\frac{90}{190}} \\ \frac{96}{271} =\ & 0.3542 \leq \Pr(y \in B \mid x = x_0) \leq 0.9852 = \frac{267}{271} \end{aligned}$$

Compare this with the identification region when only outcomes are missing.

$$\frac{96}{190} = 0.5053 \leq \Pr(y \in B \mid x = x_0) \leq 0.9789 = \frac{186}{190}$$

Clearly, the latter identification region is narrower than the former.

Example 18 (uninformative evidence) *Suppose the sampling process is altered so that* $\Pr(y \in B \mid x = x_0, z = 1)$ *is unobserved also* $\Pr(z){=}\frac{100}{190}$

and $\Pr(x = x_0 \mid z = 1) = 0$*. The identification region based on the data when outcomes and covariates are jointly missing is*

$$\begin{aligned}
\frac{0\frac{100}{190}}{0\frac{100}{190}+\frac{90}{190}} &\leq \Pr(y \in B \mid x = x_0) \leq \frac{0\frac{100}{190}}{0\frac{100}{190}+\frac{90}{190}} + \frac{\frac{90}{190}}{0\frac{100}{190}+\frac{90}{190}} \\
0 &\leq \Pr(y \in B \mid x = x_0) \leq 1
\end{aligned}$$

The sampling process is uninformative for $\Pr(y \in B \mid x = x_0)$*.*

11.1.19 Missing at random versus missing by choice

Missing at random is a common, yet controversial, identifying condition. Missing at random makes the analysis straightforward as the observed data are interpreted to be informative of $y \mid x$. Hence,

$$\Pr(y \mid x, z = 0) = \Pr(y \mid x, z = 1)$$

This invariance condition is often questioned, especially when potential outcomes are missing by choice.

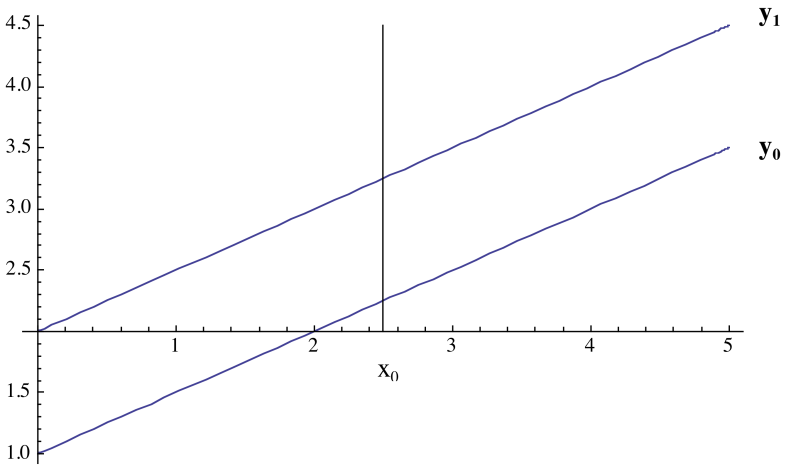
Much of the controversy regarding missing at random stems from arguments that data are missing by choice. For example, in the study of wages, it is common to invoke reservation wage homogeneity. Wage outcomes are missing for individuals out of the workforce since their wage opportunities are below the reservation wage while those employed earn wages above the reservation wage — hence, outcomes are missing by choice. Reservation wage homogeneity, or more generally, missing by choice is inconsistent with missing at random but these assertions are nonrefutable.

11.1.20 Stochastic dominance and treatment effects

If credible, stochastic dominance may help the analyst identify treatment effects amidst missing potential outcomes by choice. We next briefly consider two cases: homogeneous treatment effects and heterogeneous treatment effects. Then, we review various instrumental variable strategies for partially identifying missing outcomes in the selection problem.

11.1.21 Examples

Example 19 (stoch. dominance and treatment effect homogeneity) *Suppose binary treatment selection (*$D = 1$ *treated and* $D = 0$ *untreated) respects uniformity (that is, the propensity to adopt treatment is monotonic in the covariates) such that everyone below some threshold* x_0 *adopts no* *treatment and everyone above the threshold adopts treatment.*



homogeneous treatment effect

The treatment effect is the difference between potential outcomes with treatment , Y_1, and potential outcomes without treatment, Y_0.

$$TE = Y_1 - Y_0$$

But, an individual's treatment effect is never observed as an individual either adopts treatment or not, so the analyst looks to population level parameters (typically, means and quantiles). If Y_1 first order stochastically dominates Y_0 (or the reverse), then treatment effect homogeneity may be supported. If so, parameters respecting stochastic dominance (means and quantiles) involve the same treatment effect. For instance, $ATE(x) = ATT(x) = ATUT(x) = MedTE(x)$ where

$$\begin{aligned}
ATE(x) &= E[Y_1 - Y_0 \mid x] \\
ATT(x) &= E[Y_1 - Y_0 \mid x, D = 1] \\
ATUT(x) &= E[Y_1 - Y_0 \mid x, D = 0] \\
MedTE(x) &= Median[Y_1 \mid x] - Median[Y_0 \mid x]
\end{aligned}$$

This is a form of missing by choice. The counterfactuals, $(Y_1 \mid x, D = 0) = (Y_1 \mid x, z = 0)$ *and* $(Y_0 \mid x, D = 1) = (Y_0 \mid x, z = 0)$*, are missing potential outcome data. Homogeneity entails, say,*

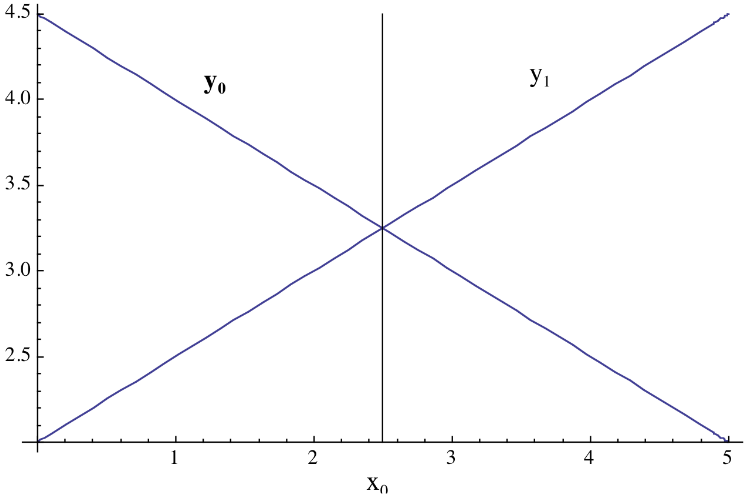
$$\begin{aligned} ATT\,(x) &= E\,[Y_1 - Y_0 \mid x, D = 1] > 0 \\ E\,[Y_1 \mid x, z = 1] &> E\,[Y_0 \mid x, z = 0] \end{aligned}$$

as well as

$$\begin{aligned} ATUT\,(x) &= E\,[Y_1 - Y_0 \mid x, D = 0] > 0 \\ E\,[Y_1 \mid x, z = 0] &> E\,[Y_0 \mid x, z = 1] \end{aligned}$$

Therefore, missing potential outcomes stochastically dominate observed outcomes to the left of x_0 *and observed outcomes stochastically dominate missing potential outcomes to the right of the threshold* x_0*.*

Example 20 (stochastic dominance and self-selection) *Suppose treatment adoption respects uniformity (as in example 19) and observed outcomes stochastically dominate missing potential outcomes.*



heterogeneous treatment effect

That is,

$$\begin{array}{rcl}(Y_1 \mid x, z=1) & = & (Y_1 \mid x, D=1) \\ & > & (Y_0 \mid x, z=0) = (Y_0 \mid x, D=1)\end{array}$$

and

$$\begin{array}{rcl}(Y_0 \mid x, z=1) & = & (Y_0 \mid x, D=0) \\ & > & (Y_1 \mid x, z=0) = (Y_1 \mid x, D=0)\end{array}$$

Outcomes are clearly heterogeneous and, in fact, support a strong form of missing by choice — individual's self-select according to the most beneficial choice (such as described by the Roy model).

$$\begin{array}{rcl}ATT(x) & = & E[Y_1 - Y_0 \mid x, D=1] > 0 \\ E[Y_1 \mid x, z=1] & > & E[Y_0 \mid x, z=0]\end{array}$$

and

$$\begin{array}{rcl}ATUT(x) & = & E[Y_1 - Y_0 \mid x, D=0] < 0 \\ E[Y_1 \mid x, z=0] & < & E[Y_0 \mid x, z=1]\end{array}$$

Thus, the implications of this form of stochastic dominance are quite different from example 19.

Next, we discuss the selection problem in more detail along with a variety of weaker stochastic dominance or monotonicity identifying conditions.

11.2 Selection problem

The foregoing discussions of missing outcome data apply to the selection problem as indicated in the treatment effect examples 19 and 20. A key characteristic of the selection problem is there may be a link between outcomes across treatments. That is, observed outcomes may help identify counterfactual potential outcomes. Also, the analyst may have reservations regarding noncompliance or partial compliance when treatment is assigned. In accounting, earnings management as an equilibrium reporting strategy suggests partial compliance (compliance only within auditor-vetted parameters) is common place.

Suppose we're interested in the impact of (perhaps, assigned) treatment on the entire population. Due to the counterfactual nature of the question, this cannot be addressed from the data alone at the individual level. However, various complementary conditions yield point or partial identification strategies for population-level parameters, D. The treatment effect is $TE =$

$y(t) - y(s)$ where $t \geq s$. Accordingly, treatment effect parameters (that respect stochastic dominance) may be defined as either $D[y(t)] - D[y(s)]$, which summarizes outcome associated with each treatment level, or focus on the incremental impact of treatment, $D[y(t) - y(s)]$. The latter is more common in the extant literature but either may be the quantity of interest to the analyst. When the focus is on expectations the two treatment effect parameters are the same $E[y(t)] - E[y(s)] = E[y(t) - y(s)]$, but not necessarily for quantiles, $Q_\alpha[y(t)] - Q_\alpha[y(s)] \neq Q_\alpha[y(t) - y(s)]$.[12] Unless otherwise noted, we focus on expectations and next remind ourselves of the impact of missing outcome data on the mean.

Suppose outcomes are bounded, $y \in [y_0, y_1]$, covariates are denoted x, potential treatment is denoted t, treatment selected is denoted z, and instruments are denoted ν, then $y(t)$ is potential outcome with treatment t and y is observed outcome. The counterfactual nature of the selection problem combined with the law of iterated expectations gives

$$\begin{aligned} E[y(t) \mid x = x] &= E[y \mid x = x, z = t] \Pr(z = t \mid x = x) \\ &\quad + E[y(t) \mid x = x, z \neq t] \Pr(z \neq t \mid x = x) \end{aligned}$$

The evidence is informative about $E[y(t) \mid x = x, z = t]$ and $\Pr(z \mid x = x)$ but uninformative of $E[y(t) \mid x = x, z \neq t]$. Hence, the identification region based on the data alone is

$$\begin{aligned} & E[y \mid x = x, z = t] \Pr(z = t \mid x = x) + y_0 \Pr(z \neq t \mid x = x) \\ \leq\ & E[y(t) \mid x = x] \leq \\ & E[y \mid x = x, z = t] \Pr(z = t \mid x = x) + y_1 \Pr(z \neq t \mid x = x) \end{aligned}$$

Then, the identification region for the conditional average treatment effect is

$$\begin{aligned} & E[y \mid x = x, z = t_2] \Pr(z = t_2 \mid x = x) + y_0 \Pr(z \neq t_2 \mid x = x) \\ & - \{E[y \mid x = x, z = t_1] \Pr(z = t_1 \mid x = x) + y_1 \Pr(z \neq t_1 \mid x = x)\} \\ \leq\ & E[y(t_2) - y(t_1) \mid x = x] \leq \\ & E[y \mid x = x, z = t_2] \Pr(z = t_2 \mid x = x) + y_1 \Pr(z \neq t_2 \mid x = x) \\ & - \{E[y \mid x = x, z = t_1] \Pr(z = t_1 \mid x = x) + y_0 \Pr(z \neq t_1 \mid x = x)\} \end{aligned}$$

That is, the lower bound for the average treatment effect involves the minimum value of $y(t_2)$ less the maximum value of $y(t_1)$. Similarly, the upper bound for the average treatment effect involves the maximum value of $y(t_2)$ less the minimum value of $y(t_1)$.

[12] When we introduce monotone treatment response (MTR) in section 11.3.6, we discuss MTR-identification regions for the two treatment effect parameters.

11.2.1 Partial compliance with random assignment

To address partial compliance, we define assignment by ζ and continue to denote treatment adopted by z with other details as immediately above. First, consider full compliance with random assignment which leads to point identification of $E\left[y\left(t\right) \mid x\right]$. Random assignment

$$\Pr\left(y\left(t\right) \mid x,\zeta\right) = \Pr\left(y\left(t\right) \mid x\right)$$

combined with full compliance gives

$$E\left[y\left(t\right) \mid x\right] = E\left[y\left(t\right) \mid x,\zeta = t\right]$$

Iterated expectations produces

$$\begin{aligned} E\left[y\left(t\right) \mid x,\zeta = t\right] &= E\left[y\left(t\right) \mid x,\zeta = t,z = t\right]\Pr\left(z = t \mid x,\zeta = t\right) \\ &\quad +E\left[y\left(t\right) \mid x,\zeta = t,z \neq t\right]\Pr\left(z \neq t \mid x,\zeta = t\right) \end{aligned}$$

but $\Pr\left(z \neq t \mid x,\zeta = t\right) = 0$ so

$$E\left[y\left(t\right) \mid x,\zeta = t\right] = E\left[y\left(t\right) \mid x,\zeta = t,z = t\right]$$

and observability of $y\left(t\right)$ leads to point identification of all treatments t.

$$E\left[y \mid x,\zeta = t\right] = E\left[y\left(t\right) \mid x,\zeta = t\right] = E\left[y\left(t\right) \mid x\right]$$

Now, consider partial identification. First, we consider partial identification of $\Pr\left(y\left(t\right) \mid x\right)$ where $\Pr\left(y\left(t\right) \mid x,\zeta\right) = \Pr\left(y\left(t\right) \mid x\right)$, then we consider partial identification of $E\left[y\left(t\right) \mid x\right]$. By the law of total probability for any $t^{'}$

$$\begin{aligned} \Pr\left(y\left(t\right) \mid x,\zeta = t^{'}\right) &= \Pr\left(y\left(t\right) \mid x,\zeta = t^{'},z = t\right)\Pr\left(z = t \mid x,\zeta = t^{'}\right) \\ &\quad +\Pr\left(y\left(t\right) \mid x,\zeta = t^{'},z \neq t\right)\Pr\left(z \neq t \mid x,\zeta = t^{'}\right) \end{aligned}$$

Then, observability gives the identification region for $\Pr\left(y\left(t\right) \mid x,\zeta = t^{'}\right)$

$$\begin{aligned} & \Pr\left(y\left(t\right) \mid x,\zeta = t^{'},z = t\right)\Pr\left(z = t \mid x,\zeta = t^{'}\right) \\ & +0\Pr\left(z \neq t \mid x,\zeta = t^{'}\right) \\ \leq\ & \Pr\left(y\left(t\right) \mid x,\zeta = t^{'}\right) \leq \\ & \Pr\left(y\left(t\right) \mid x,\zeta = t^{'},z = t\right)\Pr\left(z = t \mid x,\zeta = t^{'}\right) \\ & +1\Pr\left(z \neq t \mid x,\zeta = t^{'}\right) \end{aligned}$$

or

$$\Pr\left(y\left(t\right) \mid x,\zeta = t^{'},z = t\right)\Pr\left(z = t \mid x,\zeta = t^{'}\right) + \gamma_{t'}\Pr\left(z \neq t \mid x,\zeta = t^{'}\right)$$

where $\gamma_{t'}$ is any feasible value of $\Pr\left(y(t) \mid x, \zeta = t^{'}, z \neq t\right)$. Randomization says $\Pr(y(t) \mid x, \zeta) = \Pr(y(t) \mid x)$, in other words, any distribution that lies within the $t^{'}$ regions is a feasible distribution for $\Pr(y(t) \mid x)$. Hence, the identification region for $\Pr(y(t) \mid x)$ is the intersection of the $t^{'}$ regions.

$$\cap_{t'} \left\{ \begin{array}{c} \Pr\left(y(t) \mid x, \zeta = t^{'}, z = t\right) \Pr\left(z = t \mid x, \zeta = t^{'}\right) \\ +\gamma_{t'} \Pr\left(z \neq t \mid x, \zeta = t^{'}\right) \end{array} \right\}$$

Similarly, partial compliance leads to partial identification of $E[y(t) \mid x]$ based on random assignment. Random assignment implies

$$E[y(t) \mid x] = E[y(t) \mid x, \zeta]$$

Iterated expectations produces

$$\begin{aligned} & E[y(t) \mid x] \\ = \; & \Pr(\zeta = t \mid x) \{E[y(t) \mid x, \zeta = t, z = t] \Pr(z = t \mid x, \zeta = t) \\ & +E[y(t) \mid x, \zeta = t, z \neq t] \Pr(z \neq t \mid x, \zeta = t)\} \\ & +\Pr(\zeta \neq t \mid x) \{E[y(t) \mid x, \zeta \neq t, z = t] \Pr(z = t \mid x, \zeta \neq t) \\ & +E[y(t) \mid x, \zeta \neq t, z \neq t] \Pr(z \neq t \mid x, \zeta \neq t)\} \end{aligned}$$

Observability leads to

$$\begin{aligned} & E[y(t) \mid x, \zeta = t] \\ = \; & \Pr(\zeta = t \mid x) \{E[y \mid x, \zeta = t, z = t] \Pr(z = t \mid x, \zeta = t) \\ & +E[y(t) \mid x, \zeta = t, z \neq t] \Pr(z \neq t \mid x, \zeta = t)\} \\ & +\Pr(\zeta \neq t \mid x) \{E[y \mid x, \zeta \neq t, z = t] \Pr(z = t \mid x, \zeta \neq t) \\ & +E[y(t) \mid x, \zeta \neq t, z \neq t] \Pr(z \neq t \mid x, \zeta \neq t)\} \end{aligned}$$

Random assignment $\Pr(y(t) \mid x, \zeta) = \Pr(y(t) \mid x)$ leads to

$$\begin{array}{ccc} E[y(t) \mid x, \zeta = t] & = & E[y(t) \mid x, \zeta \neq t] \\ \begin{array}{c} E[y(t) \mid x, \zeta = t, z = t] \\ \times \Pr(z = t \mid x, \zeta = t) \\ +E[y(t) \mid x, \zeta = t, z \neq t] \\ \times \Pr(z \neq t \mid x, \zeta = t) \end{array} & = & \begin{array}{c} E[y(t) \mid x, \zeta \neq t, z = t] \\ \times \Pr(z = t \mid x, \zeta \neq t) \\ +E[y(t) \mid x, \zeta \neq t, z \neq t] \\ \times \Pr(z \neq t \mid x, \zeta \neq t) \end{array} \end{array}$$

Since $E[y(t) \mid x, \zeta, z = t]$ for all ζ and $\Pr(z \mid x, \zeta)$ are provided by the empirical evidence but $E[y(t) \mid x, \zeta, z \neq t]$ for all ζ are not, the identification region for any assignment $\zeta = t^{'}$ is

$$\begin{aligned} & E\left[y \mid x, \zeta = t^{'}, z = t\right] \Pr\left(z = t \mid x, \zeta = t^{'}\right) + y_0 \Pr\left(z \neq t \mid x, \zeta = t^{'}\right) \\ \leq \; & E\left[y(t) \mid x, \zeta = t^{'}\right] \leq \\ & E\left[y \mid x, \zeta = t^{'}, z = t\right] \Pr\left(z = t \mid x, \zeta = t^{'}\right) + y_1 \Pr\left(z \neq t \mid x, \zeta = t^{'}\right) \end{aligned}$$

Hence, the identification region for $E[y(t) \mid x]$ is the intersection of all assignments as their expectations are equal given random assignment.

$$\begin{aligned}
&\cap_{t'}\{[E\left[y \mid x, \zeta = t^{'}, z = t\right]\Pr\left(z = t \mid x, \zeta = t^{'}\right) \\
&+y_0 \Pr\left(z \neq t \mid x, \zeta = t^{'}\right), \\
&E\left[y \mid x, \zeta = t^{'}, z = t\right]\Pr\left(z = t \mid x, \zeta = t^{'}\right) \\
&+y_1 \Pr\left(z \neq t \mid x, \zeta = t^{'}\right)]\}
\end{aligned}$$

11.2.2 Example

Example 21 (compliance and random assignment) *Suppose the DGP is*

$y(1)$	$y(0)$	$TE = y(1) - y(0)$	y_{fc}	z_{fc}	y_{pc}	z_{pc}	ζ
0	0	0	0	1	0	0	1
0.25	-0.25	0.5	0.25	1	0.25	1	1
0.5	-0.5	1	0.5	1	0.5	1	1
0.75	-0.75	1.5	0.75	1	0.75	1	1
1	-1	2	1	1	-1	0	1
0	0	0	0	0	0	0	0
0.25	-0.25	0.5	-0.25	0	-0.25	0	0
0.5	-0.5	1	-0.5	0	-0.5	0	0
0.75	-0.75	1.5	-0.75	0	-0.75	0	0
1	-1	2	-1	0	1	1	0

where y_{fc} and z_{fc} are outcome and treatment adopted with full compliance and y_{pc} and z_{pc} are outcome and treatment adopted with partial compliance. Identification bounds for outcome and treatment effect means given full *compliance and partial compliance with random assignment* ζ *are*

lower bound	*parameter*	*upper bound*
0.5	$E[y_{fc}(1)]$	0.5
−0.5	$E[y_{fc}(0)]$	−0.5
1.0	$E[y_{fc}(1) - y_{fc}(0)]$	1.0

(a) full compliance with random assignment

lower bound	*parameter*	*upper bound*
−0.1	$E[y_{pc}(1) \mid \zeta = 1]$	0.7
−0.6	$E[y_{pc}(1) \mid \zeta = 0]$	1.0
−0.8	$E[y_{pc}(0) \mid \zeta = 1]$	0.4
−0.5	$E[y_{pc}(0) \mid \zeta = 0]$	−0.1
−0.1	$E[y_{pc}(1)]$	0.7
−0.5	$E[y_{pc}(0)]$	−0.1
0.0	$E[y_{pc}(1) - y_{pc}(0)]$	1.2

(b) partial compliance with random assignment

11.2.3 Partial compliance with nonrandom assignment

Now, consider nonrandom assignment. First, full compliance with nonrandom assignment (or based on the data alone) involves

$$\begin{aligned}
& E[y(t) \mid x] \\
= \; & \Pr(\zeta = t \mid x) E[y(t) \mid x, \zeta = t] \\
& + \Pr(\zeta \neq t \mid x) E[y(t) \mid x, \zeta \neq t] \\
= \; & \Pr(\zeta = t \mid x) \{\Pr(z = t \mid x, \zeta = t) E[y(t) \mid x, \zeta = t, z = t] \\
& + \Pr(z \neq t \mid x, \zeta = t) E[y(t) \mid x, \zeta = t, z \neq t]\} \\
& + \Pr(\zeta \neq t \mid x) \{\Pr(z = t \mid x, \zeta \neq t) E[y(t) \mid x, \zeta \neq t, z = t] \\
& + \Pr(z \neq t \mid x, \zeta \neq t) E[y(t) \mid x, \zeta \neq t, z \neq t]\}
\end{aligned}$$

Full compliance implies $\Pr(z = t \mid x, \zeta \neq t) = \Pr(z \neq t \mid x, \zeta = t) = 0$ thus the above expression simplifies to

$$\begin{aligned}
& E[y(t) \mid x] \\
= \; & \Pr(\zeta = t \mid x) \Pr(z = t \mid x, \zeta = t) E[y(t) \mid x, \zeta = t, z = t] \\
& + \Pr(\zeta \neq t \mid x) \Pr(z \neq t \mid x, \zeta \neq t) E[y(t) \mid x, \zeta \neq t, z \neq t]
\end{aligned}$$

Since everything in this expression but $E[y(t) \mid x, \zeta \neq t, z \neq t]$ is observable from the empirical evidence, the identification region for nonrandom assignment of any treatment t with full compliance is

$$\begin{aligned}
&\Pr(\zeta = t \mid x)\Pr(z = t \mid x, \zeta = t)\, E[y \mid x, \zeta = t, z = t] \\
&+\Pr(\zeta \neq t \mid x)\Pr(z \neq t \mid x, \zeta \neq t)\, y_0 \\
\leq\ & E[y(t) \mid x] \leq \\
&\Pr(\zeta = t \mid x)\Pr(z = t \mid x, \zeta = t)\, E[y \mid x, \zeta = t, z = t] \\
&+\Pr(\zeta \neq t \mid x)\Pr(z \neq t \mid x, \zeta \neq t)\, y_1
\end{aligned}$$

Next, we consider partial compliance with nonrandom assignment. Partial compliance with nonrandom assignment leads to

$$\begin{aligned}
& E[y(t) \mid x] \\
=\ & \Pr(\zeta = t \mid x)\{\Pr(z = t \mid x, \zeta = t)\, E[y(t) \mid x, \zeta = t, z = t] \\
&+\Pr(z \neq t \mid x, \zeta = t)\, E[y(t) \mid x, \zeta = t, z \neq t]\} \\
&+\Pr(\zeta \neq t \mid x)\{\Pr(z = t \mid x, \zeta \neq t)\, E[y(t) \mid x, \zeta \neq t, z = t] \\
&+\Pr(z \neq t \mid x, \zeta \neq t)\, E[y(t) \mid x, \zeta \neq t, z \neq t]\}
\end{aligned}$$

Everything but $E[y(t) \mid x, \zeta \neq t, z \neq t]$ and $E[y(t) \mid x, \zeta = t, z \neq t]$ is observable from the empirical evidence, the identification region for nonrandom assignment (in other words, the data alone) of any treatment t with partial compliance is

$$\begin{aligned}
&\Pr(\zeta = t \mid x)\{\Pr(z = t \mid x, \zeta = t)\, E[y \mid x, \zeta = t, z = t] \\
&+\Pr(z \neq t \mid x, \zeta = t)\, y_0\} \\
&+\Pr(\zeta \neq t \mid x)\{\Pr(z = t \mid x, \zeta \neq t)\, E[y \mid x, \zeta \neq t, z = t] \\
&+\Pr(z \neq t \mid x, \zeta \neq t)\, y_0\} \\
\leq\ & E[y(t) \mid x] \leq \\
&\Pr(\zeta = t \mid x)\{\Pr(z = t \mid x, \zeta = t)\, E[y \mid x, \zeta = t, z = t] \\
&+\Pr(z \neq t \mid x, \zeta = t)\, y_1\} \\
&+\Pr(\zeta \neq t \mid x)\{\Pr(z = t \mid x, \zeta \neq t)\, E[y \mid x, \zeta \neq t, z = t] \\
&+\Pr(z \neq t \mid x, \zeta \neq t)\, y_1\}
\end{aligned}$$

11.2.4 Example

Example 22 (compliance and nonrandom assignment) *Suppose the DGP is a nonrandom assignment variation of example 21*

$y(1)$	$y(0)$	$TE = y(1) - y(0)$	y_{fc}	z_{fc}	y_{pc}	z_{pc}	ζ
0	0	0	0	0	0	0	0
0.25	-0.25	0.5	-0.25	0	0.25	1	0
0.5	-0.5	1	0.5	1	0.5	1	1
0.75	-0.75	1.5	0.75	1	0.75	1	1
1	-1	2	1	1	-1	0	1
0	0	0	0	0	0	0	0
0.25	-0.25	0.5	-0.25	0	-0.25	0	0
0.5	-0.5	1	-0.5	0	-0.5	0	0
0.75	-0.75	1.5	0.75	1	-0.75	0	1
1	-1	2	1	1	1	1	1

Identification bounds for outcome and treatment effect means given full compliance and partial compliance with nonrandom assignment ζ are

lower bound	*parameter*	*upper bound*
-0.1	$E[y_{fc}(1)]$	0.9
-0.6	$E[y_{fc}(0)]$	0.4
-0.5	$E[y_{fc}(1) - y_{fc}(0)]$	1.5

(a) full compliance with nonrandom assignment

lower bound	*parameter*	*upper bound*
0.05	$E[y_{pc}(1) \mid \zeta = 1]$	0.85
-0.75	$E[y_{pc}(1) \mid \zeta = 0]$	0.85
-0.95	$E[y_{pc}(0) \mid \zeta = 1]$	0.25
-0.35	$E[y_{pc}(0) \mid \zeta = 0]$	0.05
-0.35	$E[y_{pc}(1)]$	0.85
-0.65	$E[y_{pc}(0)]$	0.15
-0.5	$E[y_{pc}(1) - y_{pc}(0)]$	1.5

(b) partial compliance with nonrandom assignment

Not surprisingly, these bounds are wider than those based on random assignment as random assignment involves the intersection of the assignment bounds while nonrandom involves their probability-weighted average. In other words, random assignment has identifying power.

11.2.5 Various instrumental variable strategies

A variety of instrumental variable definitions have been proposed in the literature. Below we review some common definitions and their implications for the selection problem.

11.2.6 Outcomes missing-at-random (MAR)

Outcomes missing at random (*MAR*) imply

$$\Pr\left(y\left(t\right)\mid x=x,\nu\right)=\Pr\left(y\left(t\right)\mid x=x,\nu,z=t\right)=\Pr\left(y\left(t\right)\mid x=x,\nu,z\neq t\right)$$

Combining the law of total probability with *MAR*, $\Pr\left(y\left(t\right)\mid x=x\right)$ is point-identified.

$$\begin{aligned}\Pr\left(y\left(t\right)\mid x=x\right) &= \sum_{v}\Pr\left(y\left(t\right)\mid x=x,\nu=v\right)\Pr\left(\nu=v\mid x=x\right)\\ &= \sum_{v}\Pr\left(y\mid x=x,\nu=v,z=t\right)\Pr\left(\nu=v\mid x=x\right)\\ &= \sum_{v}\Pr\left(y\left(t\right)\mid x=x,\nu=v,z\neq t\right)\Pr\left(\nu=v\mid x=x\right)\end{aligned}$$

Hence, *MAR* point identifies average and quantile treatment effects.

$$D\left[y\left(t_2\right)\mid x\right]-D\left[y\left(t_1\right)\mid x\right]$$

11.2.7 Statistical independence of outcomes and instruments (SI)

Statistical independence of outcomes and instruments (*SI*) implies

$$\Pr\left(y\left(t\right)\mid x=x,\nu\right)=\Pr\left(y\left(t\right)\mid x=x\right)$$

Combining the law of total probability with *SI*, the identification region for $\Pr\left(y\left(t\right)\mid x=x\right)$ is the narrowest or intersection region for $v\in V$

$$\begin{aligned}&\cap_{v\in V}\{[\Pr\left(y\mid x=x,\nu=v,z=t\right)\Pr\left(z=t\mid x=x,\nu=v\right),\\ &\Pr\left(y\mid x=x,\nu=v,z=t\right)\Pr\left(z=t\mid x=x,\nu=v\right)\\ &+\Pr\left(z\neq t\mid x=x,\nu=v\right)]\}\end{aligned}$$

Notice the similarity to the earlier discussion of partial compliance with random assignment. In that setting, assignment ζ serves as an instrument. This identification region is a bit abstract, another example helps clarify the identifying power of *SI* is a product of variation in the instrument ν.

11.2.8 Examples

Example 23 (*SI* identifying power through variation in ν) *Suppose the DGP is as follows*

y	$y(1)$	$y(0)$	$TE = y(1) - y(0)$	x	z	v
2	2	1	1	1	1	0
4	4	2	2	2	1	0
6	6	3	3	3	1	1
8	8	4	4	4	1	1
1	2	1	1	1	0	0
2	4	2	2	2	0	0
3	6	3	3	3	0	0
4	8	4	4	4	0	1

The identification bounds based on the data alone are

$$\begin{aligned} & \Pr(y \mid x = x, \nu = v, z = t) \Pr(z = t \mid x = x, \nu = v) \\ \leq \; & \Pr(y(t) \mid x) \leq \\ & \Pr(y \mid x = x, \nu = v, z = t) \Pr(z = t \mid x = x, \nu = v) \\ & + \Pr(z \neq t \mid x = x, \nu = v) \end{aligned}$$

In other words,

$$\begin{array}{ll} \frac{1}{2} \leq \Pr(y(1) = 2 \mid x = 1) \leq 1 & 0 \leq \Pr(y(1) \neq 2 \mid x = 1) \leq \frac{1}{2} \\ \frac{1}{2} \leq \Pr(y(1) = 4 \mid x = 2) \leq 1 & 0 \leq \Pr(y(1) \neq 4 \mid x = 2) \leq \frac{1}{2} \\ \frac{1}{2} \leq \Pr(y(1) = 6 \mid x = 3) \leq 1 & 0 \leq \Pr(y(1) \neq 6 \mid x = 3) \leq \frac{1}{2} \\ \frac{1}{2} \leq \Pr(y(1) = 8 \mid x = 4) \leq 1 & 0 \leq \Pr(y(1) \neq 8 \mid x = 4) \leq \frac{1}{2} \end{array}$$

and

$$\begin{array}{ll} \frac{1}{2} \leq \Pr(y(0) = 1 \mid x = 1) \leq 1 & 0 \leq \Pr(y(0) \neq 1 \mid x = 1) \leq \frac{1}{2} \\ \frac{1}{2} \leq \Pr(y(0) = 2 \mid x = 2) \leq 1 & 0 \leq \Pr(y(0) \neq 2 \mid x = 2) \leq \frac{1}{2} \\ \frac{1}{2} \leq \Pr(y(0) = 3 \mid x = 3) \leq 1 & 0 \leq \Pr(y(0) \neq 3 \mid x = 3) \leq \frac{1}{2} \\ \frac{1}{2} \leq \Pr(y(0) = 4 \mid x = 4) \leq 1 & 0 \leq \Pr(y(0) \neq 4 \mid x = 4) \leq \frac{1}{2} \end{array}$$

On the other hand, SI bounds are

$$\begin{aligned} & \cap_{v \in V} \{ \Pr(y \mid x = x, \nu = v, z = t) \Pr(z = t \mid x = x, \nu = v) \\ \leq \; & \Pr(y(t) \mid x = x, \nu = v) = \Pr(y(t) \mid x = x) \leq \\ & \Pr(y \mid x = x, \nu = v, z = t) \Pr(z = t \mid x = x, \nu = v) \\ & + \Pr(z \neq t \mid x = x, \nu = v) \} \end{aligned}$$

These are the same as those based on the data except when $x = 3$. *In this latter case, variation in the instrument can be exploited to narrow the* *interval. In fact,* $\Pr(y(1)=6 \mid x=3)=1$ *and* $\Pr(y(0)=3 \mid x=3)=1$*; in other words, it is point-identified. To see this, recognize*

$$\begin{aligned}\Pr(y(t) \mid x) &= \Pr(y(t) \mid x,\nu) \\ \Pr(y(t) \mid x,\nu=1) &= \Pr(y(t) \mid x,\nu=0)\end{aligned}$$

implies

$$\begin{aligned}&\Pr(y(t) \mid x,\nu=1,z=t)\Pr(z=t \mid x,\nu=1) \\ &+\Pr(y(t) \mid x,\nu=1,z\neq t)\Pr(z\neq t \mid x,\nu=1) \\ = \;&\Pr(y(t) \mid x,\nu=0,z=t)\Pr(z=t \mid x,\nu=0) \\ &+\Pr(y(t) \mid x,\nu=0,z\neq t)\Pr(z\neq t \mid x,\nu=0)\end{aligned}$$

The terms involving $(y(t) \mid z\neq t)$ *are counterfactual or missing potential outcome data.*

$$\Pr(y(1)=6 \mid x=3,\nu=1,z=1)\Pr(z=1 \mid x=3,\nu=1)=1$$

implies the counterfactual

$$\Pr(y(1)=6 \mid x=3,\nu=1,z=0)\Pr(z=0 \mid x=3,\nu=1)=0$$

and $\Pr(y(1)=6 \mid x=3,\nu=1)=1$*. But, SI indicates*

$$\Pr(y(1)=6 \mid x=3,\nu=0)=\Pr(y(1)=6 \mid x=3,\nu=1)=1$$

This along with the potentially observable but unobserved

$$\Pr(y(1)=6 \mid x=3,\nu=0,z=1)\Pr(z=1 \mid x=3,\nu=0)=0$$

implies the counterfactual must be

$$\Pr(y(1)=6 \mid x=3,\nu=0,z=0)\Pr(z=0 \mid x=3,\nu=0)=1.$$

Similarly,

$$\Pr(y(0)=3 \mid x=3,\nu=0,z=0)\Pr(z=0 \mid x=3,\nu=0)=1$$

implies the counterfactual

$$\Pr(y(0)=3 \mid x=3,\nu=0,z=1)\Pr(z=1 \mid x=3,\nu=0)=0$$

and $\Pr(y(0)=3 \mid x=3,\nu=0)=1$*. But, SI indicates*

$$\Pr(y(0)=3 \mid x=3,\nu=0)=\Pr(y(0)=3 \mid x=3,\nu=1)=1$$

This along with the potentially observable but unobserved

$$\Pr(y(0)=3,z=0 \mid x=3,\nu=1)\Pr(z=0 \mid x=3,\nu=1)=0$$

implies the counterfactual must be

$$\Pr\left(y\left(0\right)=3,z=1\mid x=3,\nu=1\right)\Pr\left(z=1\mid x=3,\nu=1\right)=1.$$

That is, the intersection of the probability regions $[0,1]$ *and* $[1,1]$ *are the points*

$$\Pr\left(y\left(1\right)=6\mid x=3\right)=1$$

and

$$\Pr\left(y\left(0\right)=3\mid x=3\right)=1.$$

Hence, the conditional treatment effect $\left[y\left(1\right)-y\left(0\right)\mid x=3\right]=6-3=3$ *is identified by SI.*

Thus, combining SI with other conditions can point-identify an average treatment effect. Another example of SI point-identification involves the local average treatment effect ($LATE$) which we briefly revisit next.

11.2.9 SI point-identification of LATE

As discussed in chapter 3, a local average treatment effect for an unidentified subpopulation of compliers is point-identified via SI in combination with uniform treatment adoption for all individuals as a function of a binary instrument.

$$LATE=E\left[y\left(t_1\right)-y\left(t_0\right)\mid\nu_1\left(z=t_1\right)-\nu_0\left(z=t_0\right)=1\right]$$

where $\nu_i\left(z=t_j\right)$ is treatment adopted $z=t_0,t_1$ when the value of the instrument is $i=0,1$. Compliers adopt treatment when the instrument value is unity and no treatment otherwise.

11.2.10 Means missing-at-random (MMAR)

Means missing-at-random ($MMAR$) implies

$$E\left[y\left(t\right)\mid x=x,\nu\right]=E\left[y\left(t\right)\mid x=x,\nu,z=t\right]=E\left[y\left(t\right)\mid x=x,\nu,z\neq t\right]$$

$MMAR$ serves the mean analogous to the manner MAR serves probability assignment. Combining the law of iterated expectations with $MMAR$, $E\left[y\left(t\right)\mid x=x\right]$ is point-identified.

$$\begin{aligned}E\left[y\left(t\right)\mid x=x\right] &= \sum_v E\left[y\left(t\right)\mid x=x,\nu=v\right]\Pr\left(\nu=v\mid x=x\right)\\ &= \sum_v E\left[y\mid x=x,\nu=v,z=t\right]\Pr\left(\nu=v\mid x=x\right)\\ &= \sum_v E\left[y\left(t\right)\mid x=x,\nu=v,z\neq t\right]\Pr\left(\nu=v\mid x=x\right)\end{aligned}$$

Accordingly, $MMAR$ point identifies the conditional average treatment effect, $E\left[y\left(t_2\right)-y\left(t_1\right)\mid x=x\right]$.

11.2.11 Mean independence of outcomes and instruments (MI)

Mean independence of outcomes and instruments (*MI*) leads to

$$E\left[y\left(t\right)\mid x=x,\nu\right]=E\left[y\left(t\right)\mid x=x\right]$$

Analogous to *SI*, the *MI* identification region for $E\left[y\left(t\right)\mid x=x\right]$ is the narrowest interval for $v\in V$

$$\begin{array}{rl} & \max_{v\in V}E\left[y\cdot 1\left(z=t\right)+y_0\cdot 1\left(z\neq t\right)\mid x=x,\nu=v\right] \\ \leq & E\left[y\left(t\right)\mid x=x,\nu=v\right]=E\left[y\left(t\right)\mid x=x\right]\leq \\ & \min_{v\in V}E\left[y\cdot 1\left(z=t\right)+y_1\cdot 1\left(z\neq t\right)\mid x=x,\nu=v\right] \end{array}$$

where $1\left(\cdot\right)$ is an indicator function equal to one when the condition in parentheses is satisfied and otherwise equals zero.

11.2.12 Example

Example 24 (*MI* identifying power through variation in ν) *Return to the setup in example 23 and apply the MI condition. The identification region for expectations $E\left[y\left(t\right)\mid x=x\right]$ based on the data alone is*

$$\begin{array}{rcl} 1.5 & \leq & E\left[y\left(1\right)\mid x=1\right]\leq 5 \\ 2.5 & \leq & E\left[y\left(1\right)\mid x=2\right]\leq 6 \\ 3.5 & \leq & E\left[y\left(1\right)\mid x=3\right]\leq 7 \\ 4.5 & \leq & E\left[y\left(1\right)\mid x=4\right]\leq 8 \end{array}$$

and

$$\begin{array}{rcl} 1 & \leq & E\left[y\left(0\right)\mid x=1\right]\leq 4.5 \\ 1.5 & \leq & E\left[y\left(0\right)\mid x=2\right]\leq 5 \\ 2 & \leq & E\left[y\left(0\right)\mid x=3\right]\leq 5.5 \\ 2.5 & \leq & E\left[y\left(0\right)\mid x=4\right]\leq 6 \end{array}$$

Hence, the identification regions for the conditional average treatment effect identified by the data alone are

$$\begin{array}{rcl} -3 & \leq & E\left[y\left(1\right)-y\left(0\right)\mid x=1\right]\leq 4 \\ -2.5 & \leq & E\left[y\left(1\right)-y\left(0\right)\mid x=2\right]\leq 4.5 \\ -2 & \leq & E\left[y\left(1\right)-y\left(0\right)\mid x=3\right]\leq 5 \\ -1.5 & \leq & E\left[y\left(1\right)-y\left(0\right)\mid x=4\right]\leq 5.5 \end{array}$$

When the analyst adds the MI condition the bounds remain the same except for $x=3$ (when there is variation in the instrument ν). Then, expectations *are point-identified as the missing data have equal expectation to the observed.*

$$6 \leq E\left[y\left(1\right) \mid x=3, \nu=1\right] = E\left[y\left(1\right) \mid x=3, \nu=0\right] \leq 6$$

and

$$3 \leq E\left[y\left(0\right) \mid x=3, \nu=0\right] = E\left[y\left(0\right) \mid x=3, \nu=1\right] \leq 3$$

Thus, the conditional average treatment effect identified via MI when $x = 3$ is point identified

$$3 \leq E\left[y\left(1\right) - y\left(0\right) \mid x=3\right] \leq 3$$

while other conditional average treatment effects ($x = 1, 2, 4$) identified via MI are the same as those identified by the data as there is no meaningful variation in ν under these conditions.

11.2.13 MI point identification

When the value of the instrument is selected treatment, $\nu = z$, *MI* becomes *MMAR* — a stronger point-identifying condition.

$$E\left[y\left(t\right) \mid x=x, \nu=z\right] = E\left[y\left(t\right) \mid x=x\right]$$

Observability combined with *MI* implies

$$E\left[y\left(t\right) \mid x=x\right] = E\left[y \mid x=x, \nu=t\right]$$

and

$$E\left[y\left(t_2\right) - y\left(t_1\right) \mid x=x\right] = E\left[y \mid x=x, \nu=t_2\right] - E\left[y \mid x=x, \nu=t_1\right]$$

This case is also referred to as ignorable treatment, selection on observables, exogenous treatment selection (ETS), or unconfounded assignment.

11.2.14 Examples

Example 25 (*MI is MMAR* when $\nu = z$ point identification) *Suppose the DGP is*

$y(1)$	$y(0)$	$TE = y(1) - y(0)$	y	$\nu = z$
5	2	3	5	1
3	2	1	3	1
4	3	1	3	0
4	1	3	1	0

The average treatment effect is identified by MMAR.

$$E\left[y \mid x=x, \nu=t_2\right] - E\left[y \mid x=x, \nu=t_1\right] = 4 - 2 = 2$$

MI and outcomes linear in treatment

In addition, if the analyst combines *MI* with outcomes linear in treatment, $\Pr\left[y\left(t\right)\right]$ is point-identified. Suppose the *DGP* is

$$y_j\left(t\right) = \beta t + u_j$$

where $u_j = y_j - \beta z_j$ is the individual-specific "intercept" and the unknown slope β is common to all individuals. Mean independence[13]

$$E\left[u \mid \nu = v_0\right] = E\left[u \mid \nu = v_1\right]$$

along with meaningful variation in the instrument

$$E\left[z \mid \nu = v_0\right] \neq E\left[z \mid \nu = v_1\right]$$

implies

$$\begin{aligned} E\left[y - \beta z \mid \nu = v_0\right] &= E\left[y - \beta z \mid \nu = v_1\right] \\ E\left[y \mid \nu = v_0\right] - \beta E\left[z \mid \nu = v_0\right] &= E\left[y \mid \nu = v_1\right] - \beta E\left[z \mid \nu = v_1\right] \\ \beta &= \frac{E\left[y \mid \nu = v_0\right] - E\left[y \mid \nu = v_1\right]}{E\left[z \mid \nu = v_0\right] - E\left[z \mid \nu = v_1\right]} \end{aligned}$$

Since the data $\Pr\left(y, z, \nu\right)$ identifies $E\left[y \mid \nu\right]$ and $E\left[z \mid \nu\right]$, β, the average treatment effect, is point-identified.

Example 26 (*MI* and outcomes linear in treatment) *Suppose the DGP is*

$$y_j\left(t\right) = \beta t + u_j$$

where $\beta = ATE = 1$ *and*

$y\left(t\right)$	$z = t$	u	ν
1	1	0	1
2	2	0	1
5	3	2	1
6	4	2	1
1	3	-2	1
8	4	4	1
1	1	0	0
3	2	1	0
3	3	0	0
7	4	3	0

$\beta = \frac{E[y|\nu=v_0]-E[y|\nu=v_1]}{E[z|\nu=v_0]-E[z|\nu=v_1]} = \frac{3\frac{1}{2}-3\frac{5}{6}}{2\frac{1}{2}-2\frac{5}{6}} = 1$, *the average treatment effect, is identified via the data when combined with MI and outcomes linear in treatment.*

[13] See ch. 3 for an alternative *DGP* associated with continuous treatment where both the intercept and slope are individual specific. A point identified correlated random coefficients strategy for identifying average treatment effects is discussed.

11.3 Monotone instrumental variable strategies

Monotone instrumental variable strategies involve a weakening of the foregoing instrumental variable strategies in pursuit of increased credibility. This weakening typically takes the form of replacing equalities with weak inequalities.

11.3.1 Mean monotonicity (MM)

Mean monotonicity (*MM*) is a relaxation of *MI* where a weak inequality replaces equality. For ordered set V

$$v_2 \geq v_1 \Rightarrow \quad E\left[y\left(t\right) \mid x = x, \nu = v_2\right] \geq E\left[y\left(t\right) \mid x = x, \nu = v_1\right]$$

The identification region given *MM* and the data is

$$\begin{array}{rl} & \sum_{v \in V} \Pr\left(\nu = v\right) \left(\max_{v' \leq v} E\left[y\left(t\right) \cdot 1\left(z = t\right) + y_0 \cdot 1\left(z \neq t\right) \mid \nu = v'\right]\right) \\ \leq & E\left[y\left(t\right)\right] \leq \\ & \sum_{v \in V} \Pr\left(\nu = v\right) \left(\min_{v' \geq v} E\left[y\left(t\right) \cdot 1\left(z = t\right) + y_1 \cdot 1\left(z \neq t\right) \mid \nu = v'\right]\right) \end{array}$$

Notice if the identification region is empty then not only is *MM* refuted but also *MI* is refuted.

11.3.2 Examples

Example 27 (*MM* and variation in ν) *Suppose the DGP is*

y	$y(1)$	$y(0)$	$TE = y(1) - y(0)$	ν	z
4	4	2	2	2	1
6	6	3	3	3	1
1	2	1	1	1	0
4	8	4	4	4	0

where $y_0 = 1$ and $y_1 = 8$. The identification bounds based on the data are

$$3 \leq E\left[y\left(1\right)\right] \leq 6.5$$

$$1.75 \leq E\left[y\left(0\right)\right] \leq 5.25$$

and

$$-2.25 \leq E\left[y\left(1\right) - y\left(0\right)\right] \leq 4.75$$

while the identification bounds based on MM are considerably narrower.

$$4.25 \leq E\left[y\left(1\right)\right] \leq 5.5$$

$$1.75 \leq E\left[y\left(0\right)\right] \leq 3.25$$

and

$$1 \leq E\left[y\left(1\right) - y\left(0\right)\right] \leq 3.75$$

The DGP clearly violates MI and the MI-identification regions are empty

$$6 \leq E\left[y\left(1\right)\right] \leq 4$$

and

$$4 \leq E\left[y\left(0\right)\right] \leq 1$$

Also, the bounds for the average treatment effect refute an MI identification strategy

$$5 \leq E\left[y\left(1\right)\right] \leq 0$$

Hence, the MI strategy is refuted by the MI partial identification bounds.

Example 28 (*MM* refuted) *Suppose the DGP is slightly altered so that MM is violated.*

y	$y\left(1\right)$	$y\left(0\right)$	$TE = y\left(1\right) - y\left(0\right)$	ν	z
4	4	2	2	2	1
1.5	1.5	3	-1.5	3	1
1	2	1	1	1	0
4	8	4	4	4	0

where $y_0 = 1$*,* $y_1 = 8$*,* $E\left[y\left(1\right)\right] = 3.875$*, and* $E\left[y\left(0\right)\right] = 2.5$*. The identification bounds based on the data are*

$$1.875 \leq E\left[y\left(1\right)\right] \leq 5.375$$

$$1.75 \leq E\left[y\left(0\right)\right] \leq 5.25$$

and

$$-3.375 \leq E\left[y\left(1\right) - y\left(0\right)\right] \leq 3.625$$

The identification region based on MM is empty for $E\left[y\left(1\right)\right]$*, refuting the MM condition.*

$$3.25 \leq E\left[y\left(1\right)\right] \leq 3.125$$

$$1.75 \leq E\left[y\left(0\right)\right] \leq 3.25$$

but bounds for the average treatment effect are not empty

$$0 \leq E\left[y\left(1\right) - y\left(0\right)\right] \leq 1.375$$

The MI-identification regions are empty indicating refutation of MI as well.

$$4 \leq E\left[y\left(1\right)\right] \leq 1.5$$

$$4 \leq E\left[y\left(0\right)\right] \leq 1$$

and

$$3 \leq E\left[y\left(1\right) - y\left(0\right)\right] \leq -2.5$$

Example 29 (*MM* violated but not refuted) *Suppose the DGP is once again slightly altered so that MM is violated but the violation occurs in the missing data.*

y	$y(1)$	$y(0)$	$TE = y(1) - y(0)$	ν	z
4	4	2	2	2	1
6	6	3	3	3	1
1	2	1	1	1	0
4	2	4	−2	4	0

where $y_0 = 1$*,* $y_1 = 6$*,* $E[y(1)] = 3.5$*, and* $E[y(0)] = 2.5$*. The identification bounds based on the data are*

$$3 \leq E[y(1)] \leq 5.5$$

$$1.75 \leq E[y(0)] \leq 4.25$$

and

$$-1.25 \leq E[y(1) - y(0)] \leq 3.75$$

The MM-identification bounds are

$$4.25 \leq E[y(1)] \leq 5$$

$$1.75 \leq E[y(0)] \leq 3.25$$

and

$$1 \leq E[y(1) - y(0)] \leq 3.25$$

Even though the MM-identification region for $E[y(1)]$ *doesn't contain the mean value, the MM-identification region is not empty and consequently does not refute the MM condition. This is not surprising as the violation occurs in the missing data. As before, the MI-identification regions are empty indicating refutation of MI.*

$$6 \leq E[y(1)] \leq 4$$

$$4 \leq E[y(0)] \leq 1$$

and

$$5 \leq E[y(1) - y(0)] \leq 0$$

11.3.3 Exogenous treatment selection (ETS)

To repeat some of the foregoing discussion, exogenous treatment selection (ETS) is the most commonly employed point-identifying condition. For each treatment t

$$\begin{array}{rc} & E[y(t) \mid x = x, z = t_2] = E[y(t) \mid x = x, z = t_1] \\ \Rightarrow & E[y(t) \mid x = x] = E[y \mid x = x, z] \end{array}$$

so that

$$E[y(t_2) - y(t_1) \mid x = x] = E[y \mid x = x, z = t_2] - E[y \mid x = x, z = t_1]$$

However, *ETS* doesn't enjoy much credibility by economic analysts.

11.3.4 Monotone treatment selection (MTS)

We earlier discussed means missing monotonically (*MMM*), in the context of prediction with missing outcome data. In the context of the selection problem, monotone treatment selection (*MTS*) is more descriptive. When the value of the instrument is selected treatment, $\nu = z$, *MM* becomes *MTS* — a stronger identifying condition. *MTS* is a relaxation of *ETS* by replacing equality with inequality.

$$t_2 \geq t_1 \Rightarrow \quad E\left[y\left(t\right) \mid x = x, z = t_2\right] \geq E\left[y\left(t\right) \mid x = x, z = t_1\right]$$

The identification region based on the data and *MTS* is

$$\begin{array}{rl} & \Pr\left(z < t\right) y_0 + \Pr\left(z \geq t\right) E\left[y \mid z = t\right] \\ \leq & E\left[y\left(t\right)\right] \leq \\ & \Pr\left(z > t\right) y_1 + \Pr\left(z \leq t\right) E\left[y \mid z = t\right] \end{array}$$

MTS offers no identifying power over the data for binary treatment as the identifying power resides with the interior treatment levels.[14] However, *MTS* is refutable as illustrated below.

11.3.5 Example

Example 30 (*MTS*) *Suppose the DGP is*

y	$y(3)$	$y(2)$	$y(1)$	$y(0)$	z
0	3	2	1	0	0
2	4	3	2	1	1
4	5	4	3	2	2
6	6	5	4	3	3

where $E\left[y\left(2\right) - y\left(1\right)\right] = 1$*. The identification bounds based on the data alone are*

$$\begin{array}{rcl} 0 & \leq & E\left[y\left(0\right)\right] \leq 4.5 \\ 0.5 & \leq & E\left[y\left(1\right)\right] \leq 5 \\ 1 & \leq & E\left[y\left(2\right)\right] \leq 5.5 \\ 1.5 & \leq & E\left[y\left(3\right)\right] \leq 6 \\ -4 & \leq & E\left[y\left(2\right) - y\left(1\right)\right] \leq 5 \end{array}$$

The MTS identification bounds are the same as those based on the data alone for extreme levels of treatment $z = 0$ *and* 3 *but considerably narrower*

[14] In the binary treatment case, one can always changes the treatment labels to satisfy *MTS*.

for the other treatments.

$$\begin{aligned} 1.5 &\leq E[y(1)] \leq 4 \\ 2 &\leq E[y(2)] \leq 4.5 \\ -2 &\leq E[y(2)-y(1)] \leq 3 \end{aligned}$$

Example 31 (*MTS* refuted) *Suppose the above DGP is slightly altered*

y	$y(3)$	$y(2)$	$y(1)$	$y(0)$	z
0	3	2	1	0	0
5	4	3	5	1	1
2	5	2	3	2	2
6	6	5	4	3	3

where $E[y(2)-y(1)] = -0.25$ *which is inconsistent with MTS. The identification bounds based on the data alone are*

$$\begin{aligned} 0 &\leq E[y(0)] \leq 4.5 \\ 1.25 &\leq E[y(1)] \leq 5.75 \\ 0.5 &\leq E[y(2)] \leq 5 \\ 1.5 &\leq E[y(3)] \leq 6 \\ -5.25 &\leq E[y(2)-y(1)] \leq 3.75 \end{aligned}$$

The MTS identification bounds refute the MTS strategy in two ways. The bounds for $E[y(1)]$ *lie above those for* $E[y(2)]$ *and the bounds for the average treatment effect,* $E[y(2)-y(1)]$*, are negative (further, the treatment effect bounds don't contain the estimand).*

$$\begin{aligned} 3.75 &\leq E[y(1)] \leq 5.5 \\ 1 &\leq E[y(2)] \leq 3 \\ -4.5 &\leq E[y(2)-y(1)] \leq -0.75 \end{aligned}$$

11.3.6 Monotone treatment response (MTR)

Monotone treatment response (*MTR*) argues

$$t_2 \geq t_1 \Rightarrow \quad y_j(t_2) \geq y_j(t_1)$$

for all individuals j and all treatment pairs (t_1, t_2). *MTR* and *MTS* differ — both may apply, either may apply or neither apply.

Example 32 (Comparison of MTR-MTS conditions) *Suppose the various DGP are as exhibited in panels (a), (b), and (c).*

$y(1)$	$y(0)$	$TE = y(1) - y(0)$	z
2	1	1	1
1	0	1	0

(a) both MTR & MTS satisfied

$y(1)$	$y(0)$	$TE = y(1) - y(0)$	z
2	1	1	1
1	0	1	0
2	1	1	1
1	0	1	0
1	2	-1	1
0	1	-1	0

(b) MTS satisfied but not MTR

$y(1)$	$y(0)$	$TE = y(1) - y(0)$	z
2	1	1	1
3	2	1	0

(c) MTR satisfied but not MTS

In panel (a), MTR is satisfied as $y_1(1) > y_1(0)$ and $y_2(1) > y_2(0)$ and MTS is satisfied as $E[y(1) \mid z = 1] > E[y(1) \mid z = 0]$ and $E[y(0) \mid z = 1] > E[y(0) \mid z = 0]$. For panel (b), MTS is satisfied as $E[y(1) \mid z = 1] = \frac{5}{3} > E[y(1) \mid z = 0] = \frac{2}{3}$ as well as $E[y(0) \mid z = 1] = \frac{4}{3} > E[y(0) \mid z = 0] = \frac{1}{3}$. However, MTR is not satisfied since $y_5(1) < y_5(0)$ and $y_6(1) < y_6(0)$. Finally, panel (c) shows MTR is satisfied as $y_1(1) > y_1(0)$ and $y_2(1) > y_2(0)$ but MTS is not satisfied as $E[y(1) \mid z = 1] < E[y(1) \mid z = 0]$ and $E[y(0) \mid z = 1] < E[y(0) \mid z = 0]$.

The identification region for MTR and the data is

$$\begin{array}{ll} & E[y \cdot 1(t \geq z) + y_0 \cdot 1(t < z) \mid \nu = v] \\ \leq & E[y(t) \mid \nu = v] \leq \\ & E[y \cdot 1(t \leq z) + y_1 \cdot 1(t > z) \mid \nu = v] \end{array}$$

Further, the identification regions for the two treatment effect parameters introduced earlier are

$$0 \leq D[y(t_2)] - D[y(t_1)] \leq D[y_1(t_2)] - D[y_0(t_1)]$$

and

$$D(0) \leq D[y(t_2) - y(t_1)] \leq D[y_1(t_2) - y_0(t_1)]$$

where

$$y_{0j}(t) = \begin{array}{ll} y_j & \text{if } t \geq z_j \\ y_0 & \text{otherwise} \end{array}$$

and

$$y_{1j}(t) = \begin{array}{ll} y_j & \text{if } t \leq z_j \\ y_1 & \text{otherwise} \end{array}$$

MTR determines the lower bounds while *MTR* and the data determine the upper bounds. *MTR* bounds are typically informative even if no individual in the population under study receives treatment t. Hence, *MTR* accommodates partial prediction of outcomes associated with proposed treatments never experienced in practice.

11.3.7 Examples

Example 33 (*MTR*) *Suppose the DGP is the same as MTS example 30.*

y	$y(3)$	$y(2)$	$y(1)$	$y(0)$	$z=\nu$
0	3	2	1	0	0
2	4	3	2	1	1
4	5	4	3	2	2
6	6	5	4	3	3

The identification bounds based on the data alone are, of course, the same as before with bounds on average treatment effects

$$\begin{array}{rcl} -4 & \leq & E[y(1)-y(0)] \leq 5 \\ -4 & \leq & E[y(2)-y(1)] \leq 5 \\ -4 & \leq & E[y(3)-y(2)] \leq 5 \end{array}$$

and the MTR-identification bounds are all narrower

$$\begin{array}{rcl} 0 & \leq & E[y(0)] \leq 3 \\ 0.5 & \leq & E[y(1)] \leq 4.5 \\ 1.5 & \leq & E[y(2)] \leq 5.5 \\ 3 & \leq & E[y(3)] \leq 6 \end{array}$$

$$\begin{array}{rcl} -2.5 & \leq & E[y(1)-y(0)] \leq 4.5 \\ -3 & \leq & E[y(2)-y(1)] \leq 5 \\ -2.5 & \leq & E[y(3)-y(2)] \leq 4.5 \end{array}$$

Some of the bounds, $E[y(0)]$ *and* $E[y(3)]$*, are narrower for MTR than for MTS while* $E[y(1)]$ *and* $E[y(2)]$ *are narrower for MTS than for MTR. Recall, for MTS the average treatment effect bound is* $-2 \leq E[y(2)-y(1)] \leq 3$.

11.3.8 MTR and means of increasing functions of outcome

MTR has identifying power for means of (weakly) increasing functions of the outcome. Let $f(\cdot)$ be such a function. $E[f(y(t))]$ respects stochastic dominance and its *MTR*-identification region is

$$\begin{aligned} & f(y_0)\Pr(z>t)+E[f(y)\mid z\le t]\Pr(z\le t) \\ \le\; & E[f(y(t))]\le \\ & f(y_1)\Pr(z<t)+E[f(y)\mid z\ge t]\Pr(z\ge t) \end{aligned}$$

whereas the identification region based on the data alone is

$$\begin{aligned} & f(y_0)\Pr(z\ne t)+E[f(y)\mid z=t]\Pr(z=t) \\ \le\; & E[f(y(t))]\le \\ & f(y_1)\Pr(z\ne t)+E[f(y)\mid z=t]\Pr(z=t) \end{aligned}$$

Example 34 (*MTR* sometimes point identifies) *Suppose the DGP is a variant on the one above where* $t=0,4$, $y_0=0$, *and* $y_1=8$ *are feasible but never observed* ($z\in\{1,2,3\}$) *and we're interested in the* $E[1(y(t)\ge 2)]=\Pr(y(t)\ge 2)$.

y	$y(4)$	$y(3)$	$y(2)$	$y(1)$	$y(0)$	$z=\nu$
2	5	4	3	2	1	1
4	6	5	4	3	2	2
6	7	6	5	4	3	3

The identification bounds, including out of sample treatments, based on the data alone are

$$\begin{aligned} 0 &\le E[1(y(0)\ge 2)]=\Pr(y(0)\ge 2)\le 1 \\ \frac{1}{3} &\le E[1(y(1)\ge 2)]=\Pr(y(1)\ge 2)\le 1 \\ \frac{1}{3} &\le E[1(y(2)\ge 2)]=\Pr(y(2)\ge 2)\le 1 \\ \frac{1}{3} &\le E[1(y(3)\ge 2)]=\Pr(y(3)\ge 2)\le 1 \\ 0 &\le E[1(y(4)\ge 2)]=\Pr(y(4)\ge 2)\le 1 \end{aligned}$$

while the MTR-identification bounds can be written

$$\Pr(z\le t\cap y\ge 2)\le \Pr(y(t)\ge 2)\le \Pr(z<t\cup y\ge 2)$$

and are (weakly) narrower than those based on the data alone

$$\begin{array}{rcl}
0 & \leq & E\left[1\left(y\left(0\right)\geq 2\right)\right]=\Pr\left(y\left(0\right)\geq 2\right)\leq 1\\
\frac{1}{3} & \leq & E\left[1\left(y\left(1\right)\geq 2\right)\right]=\Pr\left(y\left(1\right)\geq 2\right)\leq 1\\
\frac{2}{3} & \leq & E\left[1\left(y\left(2\right)\geq 2\right)\right]=\Pr\left(y\left(2\right)\geq 2\right)\leq 1\\
1 & \leq & E\left[1\left(y\left(3\right)\geq 2\right)\right]=\Pr\left(y\left(3\right)\geq 2\right)\leq 1\\
1 & \leq & E\left[1\left(y\left(4\right)\geq 2\right)\right]=\Pr\left(y\left(4\right)\geq 2\right)\leq 1
\end{array}$$

Two observations are notable. First, the upper tail probability is point identified for $t = 3, 4$. Second, treatment $t = 4$ is identified via MTR even though the population under study offers no direct evidence.

11.3.9 Mean monotonicity and mean treatment response (MM-MTR)

Combining *MM* and *MTR* yields the following identification region

$$\begin{array}{rl}
& \sum_{v\in V}\Pr\left(\nu=v\right)\left(\max_{v'\leq v}E\left[y\cdot 1\left(t\geq z\right)+y_{0}\cdot 1\left(t<z\right)\mid\nu=v^{'}\right]\right)\\
\leq & E\left[y\left(t\right)\right]\leq\\
& \sum_{v\in V}\Pr\left(\nu=v\right)\left(\min_{v'\geq v}E\left[y\cdot 1\left(t\leq z\right)+y_{1}\cdot 1\left(t>z\right)\mid\nu=v^{'}\right]\right)
\end{array}$$

11.3.10 Examples

Example 35 (*MM-MTR where* $\nu = z$) *Suppose the DGP is the same as example 33. Since $\nu = z$, MM is the same as MTS and the identification bounds are the same as the MTR-MTS identification bounds (see 11.3.11 and example 38).*

$$\begin{array}{rcl}
0 & \leq & E\left[y\left(0\right)\right]\leq 3\\
1.5 & \leq & E\left[y\left(1\right)\right]\leq 3.5\\
2.5 & \leq & E\left[y\left(2\right)\right]\leq 4.5\\
3 & \leq & E\left[y\left(3\right)\right]\leq 6
\end{array}$$

$$\begin{array}{rcl}
-1.5 & \leq & E\left[y\left(1\right)-y\left(0\right)\right]\leq 3.5\\
-1 & \leq & E\left[y\left(2\right)-y\left(1\right)\right]\leq 3\\
-1.5 & \leq & E\left[y\left(3\right)-y\left(2\right)\right]\leq 3.5
\end{array}$$

Example 36 (*MM-MTR* where $\nu \neq z$) *Suppose the DGP is*

y	$y(0)$	$y(1)$	$y(2)$	z	ν
0	0	1	2	0	0
1	0	1	2	1	0
2	0	1	2	2	0
1	1	2	3	0	1
2	1	2	3	1	1
3	1	2	3	2	1
2	2	3	4	0	2
3	2	3	4	1	2
4	2	3	4	2	2

where $y_0 = 0$ and $y_1 = 4$. The identification bounds for the means based on the data are the same as those based on MM.

$$\begin{aligned}
\frac{1}{3} &\leq E[y(0)] \leq 3 \\
\frac{2}{3} &\leq E[y(1)] \leq 3\frac{1}{3} \\
1 &\leq E[y(2)] \leq 3\frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
-2\frac{1}{3} &\leq E[y(1) - y(0)] \leq 3 \\
-2\frac{1}{3} &\leq E[y(2) - y(1)] \leq 3
\end{aligned}$$

Similarly, the symmetry in the DGP leads to the same identification bounds for MTR and MM-MTR which are narrower than those based on the data or MM.

$$\begin{aligned}
\frac{1}{3} &\leq E[y(0)] \leq 2 \\
1\frac{1}{3} &\leq E[y(1)] \leq 2\frac{5}{9} \\
3 &\leq E[y(2)] \leq 3\frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
-\frac{2}{3} &\leq E[y(1) - y(0)] \leq 2\frac{2}{9} \\
\frac{4}{9} &\leq E[y(2) - y(1)] \leq 2\frac{1}{3}
\end{aligned}$$

Example 37 (*MM-MTR* with more variation in ν) *Suppose the DGP above is perturbed slightly such that only the next to last row is modified.*

y	$y(0)$	$y(1)$	$y(2)$	z	ν
0	0	1	2	0	0
1	0	1	2	1	0
2	0	1	2	2	0
1	1	2	3	0	1
2	1	2	3	1	1
3	1	2	3	2	1
2	2	3	4	0	2
4	3	4	4	1	3
4	2	3	4	2	2

where $y_0 = 0$ and $y_1 = 4$. The identification bounds for the means based on the data are

$$\begin{aligned}
\frac{1}{3} &\leq E[y(0)] \leq 3 \\
\frac{7}{9} &\leq E[y(1)] \leq 3\frac{4}{9} \\
1 &\leq E[y(2)] \leq 3\frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
-2\frac{2}{9} &\leq E[y(1) - y(0)] \leq 3\frac{1}{9} \\
-2\frac{4}{9} &\leq E[y(2) - y(1)] \leq 2\frac{8}{9}
\end{aligned}$$

while the identification bounds based on MM differ.

$$\begin{aligned}
\frac{4}{9} &\leq E[y(0)] \leq 3 \\
\frac{25}{27} &\leq E[y(1)] \leq 3\frac{4}{9} \\
1\frac{2}{9} &\leq E[y(2)] \leq 3\frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
-2\frac{2}{27} &\leq E[y(1) - y(0)] \leq 3 \\
-2\frac{2}{9} &\leq E[y(2) - y(1)] \leq 2\frac{20}{27}
\end{aligned}$$

The identification bounds for the means based on MTR differ from those above

$$\begin{aligned}\frac{1}{3} &\leq E\left[y\left(0\right)\right] \leq 2\frac{1}{9} \\ 1\frac{1}{3} &\leq E\left[y\left(1\right)\right] \leq 2\frac{2}{3} \\ 2\frac{5}{9} &\leq E\left[y\left(2\right)\right] \leq 3\frac{2}{3}\end{aligned}$$

$$\begin{aligned}-\frac{7}{9} &\leq E\left[y\left(1\right)-y\left(0\right)\right] \leq 2\frac{1}{3} \\ -\frac{1}{9} &\leq E\left[y\left(2\right)-y\left(1\right)\right] \leq 2\frac{1}{3}\end{aligned}$$

and differ from those based on MM-MTR which are the narrowest bounds.

$$\begin{aligned}\frac{4}{9} &\leq E\left[y\left(0\right)\right] \leq 2\frac{1}{9} \\ 1\frac{11}{27} &\leq E\left[y\left(1\right)\right] \leq 2\frac{2}{3} \\ 3 &\leq E\left[y\left(2\right)\right] \leq 3\frac{2}{3}\end{aligned}$$

$$\begin{aligned}-\frac{19}{27} &\leq E\left[y\left(1\right)-y\left(0\right)\right] \leq 2\frac{2}{9} \\ \frac{1}{3} &\leq E\left[y\left(2\right)-y\left(1\right)\right] \leq 2\frac{7}{27}\end{aligned}$$

As with the above monotone instrument strategies, if outcome is unbounded so is the identification region for the expected value of outcome. On the other hand, *MTR* and *MTS* combined produce a bounded identification region for the mean even if outcome is unbounded. Hence, *MTR-MTS* has considerable identifying power.

11.3.11 Mean treatment response and mean treatment selection (MTR-MTS)

Combining *MTR* and *MTS* yields the following identification region

$$\begin{aligned}&\sum_{s<t} E\left[y \mid z=s\right]\Pr\left(z=s\right)+E\left[y \mid z=t\right]\Pr\left(z\geq t\right) \\ \leq\;& E\left[y\left(t\right)\right] \leq \\ &\sum_{s>t} E\left[y \mid z=s\right]\Pr\left(z=s\right)+E\left[y \mid z=t\right]\Pr\left(z\leq t\right)\end{aligned}$$

11.3.12 Example

Example 38 (*MTR-MTS*) *Suppose the DGP is an unbounded variation of example 33.*

y	$y(3)$	$y(2)$	$y(1)$	$y(0)$	$z=\nu$
0	3	2	1	0	0
2	4	3	2	$-\infty$	1
4	∞	4	3	2	2
6	6	5	4	3	3

Even though outcome is unbounded and identification bounds for the mean are unbounded based on the data, MTR, or MTS alone, the identification bounds for MTR-MTS are the same as example 35.

$$\begin{aligned} 0 &\leq E[y(0)] \leq 3 \\ 1.5 &\leq E[y(1)] \leq 3.5 \\ 2.5 &\leq E[y(2)] \leq 4.5 \\ 3 &\leq E[y(3)] \leq 6 \end{aligned}$$

$$\begin{aligned} -1.5 &\leq E[y(1)-y(0)] \leq 3.5 \\ -1 &\leq E[y(2)-y(1)] \leq 3 \\ -1.5 &\leq E[y(3)-y(2)] \leq 3.5 \end{aligned}$$

Next, we visit quantile treatment effect (point and partial) identification strategies.

11.4 Quantile treatment effects

Quantile treatment effects (QTE) are point identified very similarly to $LATE$, local average treatment effects, when a binary instrument exists (see Abadie et al [1998]). In addition to standard identification conditions for $LATE$ (potential outcomes are independent of instruments ν, treatment z is meaningfully related to the instruments, and treatment adoption is uniformly increasing in the instruments; see chapter 3), $QTE(\theta)$ uniqueness can only be assured if the θ-quantiles for $Y(0)$ and $Y(1)$ conditional on $z_1 - z_0 = 1$ (defined below) are unique. Some explanation is in order.

Quantiles are typically defined by the distribution function,

$$F(y) = \sum_{i=y_l}^{y_u} \Pr(y_i)$$

where y_l is the lower bound of support for y and y_u is the upper bound of support for y. However, if we define another function, say

$$G(y) = \sum_{i=y_u}^{y_l} \Pr(y_i),$$

the quantile is unambiguous or unique if $F^{-1}(\theta) = G^{-1}(1-\theta)$. This statement is always true for continuous random variables but may fail for random variables with discrete support. An example helps clarify.

Suppose the data generating process is uniform$\{1, 2, 3, 4\}$. Then,

y	$F(y)$	$G(y)$
1	0.25	1.0
2	0.50	0.75
3	0.75	0.50
4	1.0	0.25

First, second, and third quartiles are ambiguous but immediate surrounding quantiles are not.

θ	$F^{-1}(\theta)$	$G^{-1}(1-\theta)$
0.24	1	1
0.25	1	2
0.26	2	2
0.49	2	2
0.5	2	3
0.51	3	3
0.74	3	3
0.75	3	4
0.76	4	4

Point identification of $QTE(\theta)$ may fail if the $\theta-$quantile for $Y(0)$ or $Y(1)$ conditional on $z_1 - z_0 = 1$ is ambiguous. When θ-quantiles for $Y(0)$ or $Y(1)$ conditional on $z_1 - z_0 = 1$ are unambiguous, the $\theta-$quantile treatment effect conditional on X is

$$\begin{aligned} QTE(\theta \mid X = x, z_1 - z_0 = 1) &\equiv Q_\theta(Y(1) \mid X = x, z_1 - z_0 = 1) \\ &\quad -Q_\theta(Y(0) \mid X = x, z_1 - z_0 = 1) \\ &= \alpha \end{aligned}$$

where $Q_\theta(\cdot)$ refers to $\theta-$quantile of the random variable, α is the conditional quantile treatment effect from a quantile regression with treatment $z = 0, 1$ and covariates X,

$$Y \equiv zY(1) + (1-z)Y(0) = \alpha z + X\beta + \varepsilon$$

$x\beta$ is the $\theta-$quantile for $Y(0)$ conditional on $X = x_i$, z_1 is treatment when the instrument $\nu = 1$ and z_0 is treatment when $\nu = 0$. Hence, $z_1 - z_0 = 1$ refers to the target subpopulation of compliers. That is, those individuals who adopt treatment when the instrument is manipulated from zero to one.

11.4.1 Identification

Given the conditions above,

$$\underset{a,b}{\arg\min} \quad E\left[\begin{array}{c}\left(1-\frac{(1-\nu)z}{\Pr(\nu=0|X)}-\frac{(1-z)\nu}{\Pr(\nu=1|X)}\right)\\ \cdot\left(Y-az-Xb\right)\cdot\left(\theta-1\left\{Y-az-Xb<0\right\}\right)\end{array}\right]$$

equals α, β. Hence, the conditional $\theta-$quantile treatment effect, α, is point identified for the (unidentified) subpopulation of compliers.

Think of this as $E[\kappa\psi]$ where conventional quantile regression is identified by

$$\underset{b}{\arg\min} \quad E[\psi]$$

that is, where the leading term in the expectation equals one.

$$\underset{b}{\arg\min} \quad E\left[(Y-Xb)\cdot\left(\theta-1\left\{Y-Xb<0\right\}\right)\right]$$

The intuition for quantile regression is that $\widehat{Y}=Xb$ is chosen to minimize the expected linear loss where the loss function is

$$C\left(\widehat{Y},Y\mid X\right)=\begin{array}{cc}c_1\left|\widehat{Y}-Y\right| & \widehat{Y}\leq Y\\ c_2\left|\widehat{Y}-Y\right| & \widehat{Y}>Y\end{array}$$

and $\theta=\frac{c_1}{c_1+c_2}$.

Now, let's explore the $\kappa=\left(1-\frac{(1-\nu)z}{\Pr(\nu=0|X)}-\frac{(1-z)\nu}{\Pr(\nu=1|X)}\right)$ term. Recall, we're interested in the quantile treatment effect for the subpopulation of compliers. In other words, the quantile treatment effect, α, conditional on $z_1>z_0$,

$$\underset{a,b}{\arg\min} \quad E\left[\psi\mid z_1>z_0\right]$$

equals α, β. $E[\psi\mid X]$ can be written

$$\begin{aligned}E[\psi\mid X] \quad = \quad & \Pr\left(z_1>z_0\mid X\right)E\left[\psi\mid X,z_1>z_0\right]\\ & +\Pr\left(z_1=z_0=1\mid X\right)E\left[\psi\mid X,z_1=z_0=1\right]\\ & +\Pr\left(z_1=z_0=0\mid X\right)E\left[\psi\mid X,z_1=z_0=0\right]\end{aligned}$$

where uniformity implies $\Pr\left(z_1<z_0\mid X\right)=0$. Rearrange to isolate the quantity of interest.

$$\begin{aligned}& E\left[\psi\mid X,z_1>z_0\right]\\ = \quad & \frac{1}{\Pr\left(z_1>z_0\mid X\right)}\\ & \cdot\left\{\begin{array}{c}E[\psi\mid X]-\Pr\left(z_1=z_0=1\mid X\right)E\left[\psi\mid X,z_1=z_0=1\right]\\ -\Pr\left(z_1=z_0=0\mid X\right)E\left[\psi\mid X,z_1=z_0=0\right]\end{array}\right\}\end{aligned}$$

Consider the subpopulation of always adopters, $z_1 = z_0 = 1$. Uniformity, $\Pr(z_1 < z_0 \mid X) = 0$, implies $z_0 = 1$ occurs only for always adopters. Hence,

$$\begin{aligned} E[\psi \mid X, z_1 = z_0 = 1] &= E[\psi \mid X, z = 1, \nu = 0] \\ &= E\left[\frac{z(1-\nu)\psi}{\Pr(\nu = 0 \mid X)\Pr(z = 1 \mid X, \nu = 0)} \mid X\right] \end{aligned}$$

and $\Pr(z = 1 \mid X, \nu = 0) = \Pr(z_1 = z_0 = 1 \mid X)$ since the leading term only occurs for the subpopulation of always adopters.

Similar, reasoning applies to the subpopulation of never adopters, $z_1 = z_0 = 0$. Uniformity, $\Pr(z_1 < z_0 \mid X) = 0$, implies $z_1 = 0$ occurs only for never adopters. Hence,

$$\begin{aligned} E[\psi \mid X, z_1 = z_0 = 0] &= E[\psi \mid X, z = 0, \nu = 1] \\ &= E\left[\frac{\nu(1-z)\psi}{\Pr(\nu = 1 \mid X)\Pr(z = 0 \mid X, \nu = 1)} \mid X\right] \end{aligned}$$

and $\Pr(z = 0 \mid X, \nu = 1) = \Pr(z_1 = z_0 = 0 \mid X)$ since the leading term only occurs for the subpopulation of never adopters.

On substitution, we have

$$\begin{aligned} & E[\psi \mid X, z_1 > z_0] \\ = \; & \frac{1}{\Pr(z_1 > z_0 \mid X)} E\left[\left(1 - \frac{z(1-\nu)}{\Pr(\nu = 0 \mid X)} - \frac{\nu(1-z)}{\Pr(\nu = 1 \mid X)}\right)\psi \mid X\right] \\ = \; & \frac{1}{\Pr(z_1 > z_0 \mid X)} E[\kappa\psi \mid X] \end{aligned}$$

Finally, as $\frac{1}{\Pr(z_1 > z_0 \mid X)}$ does not involve a or b, this term can be ignored when solving for the arguments to minimize the expectation. Integrating over X gives the identification condition.[15]

11.4.2 Estimation

Conventional quantile regression has a linear programming (LP) formulation.

$$\begin{aligned} \min_{\tau \geq 0} \quad & c^T \tau \\ s.t. \quad & A\tau = Y \end{aligned}$$

[15] The $E[k\psi]$ development is based on lemma 2 from Abadie et al [1998] and is quite general. That is, for any statistic for which a unique moment conditional on the subpopulation of complier exists, the moment can be identified via this κ-strategy. For instance, $LATE = \alpha$ is the solution $\mu, \alpha = \underset{m,a}{\arg\min} E\left[\kappa(Y - m - az)^2\right]$ where $\mu = E[Y_0 \mid z_1 > z_0]$ and $\alpha = E[Y_1 - Y_0 \mid z_1 > z_0]$.

where $c = (o, o, \theta \cdot \iota, (1-\theta) \cdot \iota)^T$, $\tau = (a^+, b^+, a^-, b^-, u^+, u^-)^T$, $u = Y - za - Xb$, $A = [z, X, -z, -X, I, -I]$, o is an $h+1$ element vector of zeroes, ι is and n element vector of ones, X is an $n \times h$ matrix of covariates, I is an $n \times n$ identity matrix, b has h elements, e^+ denotes the positive part and e^- denotes the negative part of real number e.

Estimation of $QTE(\theta)$ involves a variation on the above where c is redefined as $(o, o, \theta \cdot K, (1-\theta) \cdot K)^T$ and $K = \kappa_1, \ldots, \kappa_n$, an n element vector composed of the sample analog of $\left(1 - \frac{(1-\nu)z}{\Pr(\nu=0|X)} - \frac{(1-z)\nu}{\Pr(\nu=1|X)}\right)$. However, when κ_i is negative (for instance, $\nu = 1$ and $z = 0$) the LP is unbounded. This necessitates further modification. Two additional constraints and one additional parameter, s_i, are added for each instance where κ_i is negative.

$$\begin{array}{rcl} u_i^+ & \leq & Ms_i \\ u_i^- & \leq & M(1-s_i) \end{array}$$

where M is a large (nonbinding) constant, and $s_i \in \{0,1\}$, an integer. In other words, we now have a mixed integer linear program ($MILP$) formulation for QTE estimation. It's time for some examples.

11.4.3 Examples

The first example illustrates unconfounded quantile treatment effects as outcomes are independent of treatment. In other words, treatment serves as an instrument and the entire population is composed of compliers. As there are no covariates $X = \iota$.

Example 39 (unconfounded *QTE*) *Suppose the DGP is*

Y	$Y(1)$	$Y(0)$	$TE = Y(1) - Y(0)$	z	ν
0	2	0	2	0	0
2	2	0	2	1	1
2	4	2	2	0	0
4	4	2	2	1	1
3	5	3	2	0	0
5	5	3	2	1	1
4	6	4	2	0	0
6	6	4	2	1	1
6	8	6	2	0	0
8	8	6	2	1	1

All quantile treatment effects except $\theta = 0.2, 0.4, 0.6, 0.8$ (where $Q_\theta[Y(0)]$ and $Q_\theta[Y(1)]$ are not unique) are point identified. Some quantile treatment effects, α, along with quantiles for $Y(0)$, β, are tabulated below. Partially *identified (non-unique) quantities are indicated by intervals within which the objective function value is constant and minimized.*

θ	$\alpha = Q_\theta [Y(1)] - Q_\theta [Y(0)]$	$\beta = Q_\theta [Y(0)]$	$Q_\theta [Y(1)]$
0.1	2	0	2
0.2	$(0,4)$	$(0,2)$	$(2,4)$
0.3	2	2	4
0.4	$(1,3)$	$(2,3)$	$(4,5)$
0.5	2	3	5
0.6	$(1,3)$	$(3,4)$	$(5,6)$
0.7	2	4	6
0.8	$(0,4)$	$(4,6)$	$(6,8)$
0.9	2	6	8

Outcomes are homogeneous and since $\Pr(z_0 = 1) = 0$*, $QTE(\theta)$ for the compliers equals $QTT(\theta)$, the quantile treatment effect for the treated. Likewise, as* $\Pr(z_1 = 1) = 1$*, $QTE(\theta)$ for the compliers equals $QTUT(\theta)$, the quantile treatment effect for the untreated. This is a case of unconfounded treatment as treatment adopted serves the role of an instrument.*

Example 40 (QTE for subsample of compliers) *Suppose the DGP is a slight variation of example 39.*

Y	$Y(1)$	$Y(0)$	$TE = Y(1) - Y(0)$	z	ν
0	2	0	2	0	0
0	2	0	2	0	1
2	4	2	2	0	0
4	4	2	2	1	1
3	5	3	2	0	0
5	5	3	2	1	1
4	6	4	2	0	0
6	6	4	2	1	1
6	8	6	2	0	0
8	8	6	2	1	1

Compliers are represented by rows 3 through 10. All quantile treatment effects except $\theta = 0.25, 0.5, 0.75$ *(where* $Q_\theta [Y(0)]$ *and* $Q_\theta [Y(1)]$ *are not unique) are point identified. Some quantile treatment effects, α, along with quantiles for $Y(0)$, β, are tabulated below. Partially identified (non-unique)* *quantities are indicated by intervals.*

θ	$\alpha = Q_\theta[Y(1)] - Q_\theta[Y(0)]$	$\beta = Q_\theta[Y(0)]$	$Q_\theta[Y(1)]$
0.1	2	2	4
0.2	2	2	4
0.25	$(1,3)$	$(2,3)$	$(4,5)$
0.4	2	3	5
0.5	$(1,3)$	$(3,4)$	$(5,6)$
0.6	2	4	6
0.75	$(0,4)$	$(4,6)$	$(6,8)$
0.8	2	6	8
0.9	2	6	8

Outcomes are again homogeneous and since $\Pr(z_0 = 1) = 0$, $QTE(\theta)$ *for the compliers equals* $QTT(\theta)$, *the quantile treatment effect for the treated. Notice, even though quartiles are uniquely defined for the population that is not the case for the subpopulation of compliers.*

Example 41 (more variation in *QTE*) *Suppose the DGP involves more variation than example 40.*

Y	$Y(1)$	$Y(0)$	$TE = Y(1) - Y(0)$	z	ν
0	2	0	2	0	0
0	2	0	2	0	1
2	4	2	2	0	0
4	4	2	2	1	1
5	5	5	0	0	0
5	5	5	0	1	1
5	6	5	1	0	0
6	6	5	1	1	1
6	8	6	2	0	0
8	8	6	2	1	1

Compliers are represented by rows 3 through 10. Again, all quantile treatment effects except $\theta = 0.25, 0.5, 0.75$ *(where* $Q_\theta[Y(0)]$ *and* $Q_\theta[Y(1)]$ *are not unique) are point identified. Some quantile treatment effects,* α, *along with quantiles for* $Y(0)$, β, *are tabulated below. Partially identified (non-*

unique) quantities are indicated by intervals.

θ	$\alpha = Q_\theta[Y(1)] - Q_\theta[Y(0)]$	$\beta = Q_\theta[Y(0)]$	$Q_\theta[Y(1)]$
0.1	2	2	4
0.2	2	2	4
0.25	$(-1,3)$	$(2,5)$	$(4,5)$
0.4	0	5	5
0.5	$(0,1)$	5	$(5,6)$
0.6	1	5	6
0.75	$(0,3)$	$(5,6)$	$(6,8)$
0.8	2	6	8
0.9	2	6	8

Outcomes are heterogeneous but since $\Pr(z_0 = 1) = 0$*,* $QTE(\theta)$ *for the compliers equals* $QTT(\theta)$*, the quantile treatment effect for the treated.*

Next, we revisit monotone treatment response (*MTR*) and explore partial identification of *QTE*.

11.4.4 MTR and partial identification of QTE

MTR says if treatment $t > s$, then $y_j(t) > y_j(s)$ for all individuals j. *MTR* bounds for outcome quantity D that respects stochastic dominance are

$$D[y_0(t)] \leq D[y(t)] \leq D[y_1(t)]$$

where

$$y_{0j}(t) \equiv \begin{matrix} y_j & t \geq z_j \\ y_0 & \text{otherwise} \end{matrix}$$

$$y_{1j}(t) \equiv \begin{matrix} y_j & t \leq z_j \\ y_1 & \text{otherwise} \end{matrix}$$

and z_j is individual j's adopted treatment.

Partial identification bounds for *MTR* quantiles are

$$\begin{array}{ccc} 0 < \theta \leq \Pr(t < z) & \Rightarrow & y_0 \leq Q_\theta[y(t)] \leq Q_{\lambda_1}(y \mid t \leq z) \\ \Pr(t < z) < \theta \leq \Pr(t \leq z) & \Rightarrow & Q_{\lambda_0}(y \mid t \geq z) \leq Q_\theta[y(t)] \leq Q_{\lambda_1}(y \mid t \leq z) \\ \Pr(t \leq z) < \theta < 1 & \Rightarrow & Q_{\lambda_0}(y \mid t \geq z) \leq Q_\theta[y(t)] \leq y_1 \end{array}$$

where

$$\begin{aligned} \lambda_1 &\equiv \frac{\theta}{\Pr(t \leq z)} \\ \lambda_0 &\equiv \frac{\theta - \Pr(t < z)}{\Pr(t \geq z)} \end{aligned}$$

In the binary treatment setting, $t = 0, 1$, the *MTR* quantile bounds are

$$\begin{array}{ccc}
\begin{array}{c} 0 < \theta \leq \Pr(t < z) \\ t = 0 \end{array} & \Rightarrow & y_0 \leq Q_\theta[y(0)] \leq Q_\theta(y) \\
\\
\Pr(t < z) < \theta \leq \Pr(t \leq z) & & \\
t = 0 & \Rightarrow & Q_\theta(y \mid z = 0) \leq Q_\theta[y(0)] \leq Q_\theta(y) \\
t = 1 & \Rightarrow & Q_\theta(y) \leq Q_\theta[y(1)] \leq Q_\theta(y \mid z = 1) \\
\\
\begin{array}{c} \Pr(t \leq z) < \theta < 1 \\ t = 1 \end{array} & \Rightarrow & Q_\theta(y) \leq Q_\theta[y(1)] \leq y_1
\end{array}$$

Then, the *MTR* treatment effect for any quantity D that respects stochastic dominance (e.g.,. means or quantiles) has bounds

$$0 \leq D[y(t)] - D[y(s)] \leq D[y_1(t)] - D[y_0(s)]$$

To appreciate this result consider the bounds on the following exhaustive monotone treatment response cases.

$$\begin{array}{rclr}
s < t < z_j & \Rightarrow & y_0 \leq y_j(s) \leq y_j(t) \leq y_j & (1) \\
s < t = z_j & \Rightarrow & y_0 \leq y_j(s) \leq y_j(t) = y_j & (2) \\
s < z_j < t & \Rightarrow & y_0 \leq y_j(s) \leq y_j \leq y_j(t) \leq y_1 & (3) \\
s = z_j < t & \Rightarrow & y_j = y_j(s) \leq y_j(t) \leq y_1 & (4) \\
z_j < s < t & \Rightarrow & y_j \leq y_j(s) \leq y_j(t) \leq y_1 & (5)
\end{array}$$

For simplicity, consider the implications for quantile treatment effect bounds with binary treatment, $s = 0$ and $t = 1$. Only cases (2) and (4) apply.

Case (2) identifies quantile bounds as

$$0 < \theta \leq \Pr(t < z) \quad \Rightarrow \quad y_0 \leq Q_\theta[y(0)] \leq Q_\theta(y)$$

and

$$0 = \Pr(t < z) < \theta \leq \Pr(t \leq z) \quad \Rightarrow \quad Q_\theta(y) \leq Q_\theta[y(1)] \leq Q_\theta(y \mid z = 1)$$

Hence, the case (2) quantile treatment effect

$$QTE(\theta) = Q_\theta[y(1)] - Q_\theta[y(0)]$$

has bounds

$$0 = Q_\theta(y) - Q_\theta(y) \leq QTE(\theta) \leq Q_\theta(y \mid z = 1) - y_0$$

Case (4) identifies quantile bounds as

$$\Pr(t < z) < \theta \leq \Pr(t \leq z) \quad \Rightarrow \quad Q_\theta(y \mid z = 0) \leq Q_\theta[y(0)] \leq Q_\theta(y)$$

and

$$\Pr(t \leq z) < \theta < 1 \quad \Rightarrow \quad Q_\theta(y) \leq Q_\theta[y(1)] \leq y_1$$

Hence, the quantile treatment effect for case (4)

$$QTE(\theta) = Q_\theta[y(1)] - Q_\theta[y(0)]$$

has bounds

$$0 = Q_\theta(y) - Q_\theta(y) \leq QTE(\theta) \leq y_1 - Q_\theta(y \mid z = 0)$$

11.4.5 Quantile treatment effects based on the data alone

On the other hand, quantile treatment effects based on the data alone are wider. From section 11.1.2, the θ-quantile bounds are

$$\begin{array}{rl} & r(\theta, x) \\ \leq & Q_\theta(y \mid x) \leq \\ & s(\theta, x) \end{array}$$

where

$$r(\theta, x) = \begin{array}{cc} Q_{\frac{\theta - \Pr(\text{missing} \mid x)}{\Pr(\text{observed} \mid x)}}(y \mid x, \text{observed}) & \text{if } \Pr(\text{missing} \mid x) < \theta \\ y_0 & \text{otherwise} \end{array}$$

and

$$s(\theta, x) = \begin{array}{cc} Q_{\frac{\theta}{\Pr(\text{observed} \mid x)}}(y \mid x, \text{observed}) & \text{if } \Pr(\text{missing} \mid x) < 1 - \theta \\ y_1 & \text{otherwise} \end{array}$$

To illustrate quantile bounds for treatment effects, consider the binary treatment case. Quantile bounds based on the data alone are

$$\begin{array}{rcl} r(\theta, x, 0) & \leq & Q_\theta[y(0) \mid x] \leq s(\theta, x, 0) \\ r(\theta, x, 1) & \leq & Q_\theta[y(1) \mid x] \leq s(\theta, x, 1) \end{array}$$

so that quantile treatment effect, $Q_\theta[y(1) \mid x] - Q_\theta[y(0) \mid x]$, bounds are

$$r(\theta, x, 1) - s(\theta, x, 0) \leq QTE(\theta \mid x) \leq s(\theta, x, 1) - r(\theta, x, 0)$$

where

$$\begin{array}{rcl} r(\theta, x, 0) & = & \begin{array}{cc} Q_{\frac{\theta - \Pr(z=1)}{\Pr(z=0)}}(y \mid x, z = 0) & \text{if } \Pr(z = 1 \mid x) < \theta \\ y_0 & \text{otherwise} \end{array} \\ s(\theta, x, 0) & = & \begin{array}{cc} Q_{\frac{\theta}{\Pr(z=0)}}(y \mid x, z = 0) & \text{if } \Pr(z = 1 \mid x) < 1 - \theta \\ y_1 & \text{otherwise} \end{array} \\ r(\theta, x, 1) & = & \begin{array}{cc} Q_{\frac{\theta - \Pr(z=0)}{\Pr(z=1)}}(y \mid x, z = 1) & \text{if } \Pr(z = 0 \mid x) < \theta \\ y_0 & \text{otherwise} \end{array} \\ s(\theta, x, 1) & = & \begin{array}{cc} Q_{\frac{\theta}{\Pr(z=1)}}(y \mid x, z = 1) & \text{if } \Pr(z = 0 \mid x) < 1 - \theta \\ y_1 & \text{otherwise} \end{array} \end{array}$$

A binary treatment illustration helps illuminate some of the subtleties associated with quantile treatment effect bounds.

11.4.6 Example

Example 42 (*MTR* bounds for *QTE*) *Suppose the DGP is the same as example 40.*

Y	$Y(1)$	$Y(0)$	$TE = Y(1) - Y(0)$	z	ν
0	2	0	2	0	0
0	2	0	2	0	1
2	4	2	2	0	0
4	4	2	2	1	1
3	5	3	2	0	0
5	5	3	2	1	1
4	6	4	2	0	0
6	6	4	2	1	1
6	8	6	2	0	0
8	8	6	2	1	1

The median and first and third quartile bounds based on MTR are

θ	$Q_\theta^{lower}[y(0)]$	$Q_\theta^{upper}[y(0)]$	$Q_\theta^{lower}[y(1)]$	$Q_\theta^{upper}[y(1)]$
0.2	0	$(0,2)$	$(0,2)$	4
0.25	0	2	2	$(4,5)$
0.4	0	$(3,4)$	$(3,4)$	5
0.5	$(2,3)$	4	4	8
0.6	3	$(4,5)$	$(4,5)$	8
0.75	4	6	6	8
0.8	4	6	6	8

θ	$QTE^{lower}(\theta \mid MTR)$	$QTE^{upper}(\theta \mid MTR)$
0.2	$(0,2) - (0,2) = 0$	$4 - 0 = 4$
0.25	$2 - 2 = 0$	$5 - 0 = 5$
0.4	$(3,4) - (3,4) = 0$	$5 - 0 = 5$
0.5	$4 - 4 = 0$	$8 - 2 = 6$
0.6	$(4,5) - (4,5) = 0$	$8 - 3 = 5$
0.75	$6 - 6 = 0$	$8 - 4 = 4$
0.8	$6 - 6 = 0$	$8 - 4 = 4$

where non-unique quantiles are indicated by intervals. While these bounds may not seem very tight, MTR (in conjunction with the data) always results in informative bounds. Monotone response implies the treatment effect is never negative but the data alone may not rule out negative treatment effects. The data alone identify the following substantially wider quantile *treatment effect bounds.*

θ	$QTE^{lower}\left(\theta \mid data\right)$	$QTE^{upper}\left(\theta \mid data\right)$
0.2	$0-(0,2)=-2$	$(5,6)-0=6$
0.25	$0-2=-2$	$6-0=6$
0.4	$0-(3,4)=-4$	$8-0=8$
0.5	$0-(4,6)=-6$	$8-0=8$
0.6	$(0,4)-(6,8)=-8$	$8-(0,2)=8$
0.75	$5-8=-3$	$8-3=5$
0.8	$6-8=-2$	$8-(3,4)=5$

As indicated in example 40, treatment effects are homogeneously equal to 2 *for all unique quantiles for this DGP. The existence of a binary instrument leads to the following bounds on QTE(θ) for the subpopulation of compliers (quartile treatment effects are not point identified as quartiles are not unique for the DGP conditional on* $z_1 - z_0 = 1$*).*

θ	$QTE^{lower}\left(\theta \mid z_1 - z_0 = 1\right)$	$QTE^{upper}\left(\theta \mid z_1 - z_0 = 1\right)$
0.2	2	2
0.25	1	3
0.4	2	2
0.5	1	3
0.6	2	2
0.75	0	4
0.8	2	2

11.5 Extrapolation and the mixing problem

Suppose the analyst observes outcomes of an idealized classical, randomized experiment with full treatment compliance but wishes to extrapolate or predict outcomes in settings where treatment (or compliance) varies across individuals. That is, the classical experiment has the property $\Pr\left(y\left(t\right) \mid x, z\right) = \Pr\left(y\left(t\right) \mid x\right)$ but the analyst wishes to extrapolate to settings in which independence fails. This is referred to as the mixing problem as it seeks inference on a probability mixture from marginal probabilities.

Define a treatment policy τ as one that specifies the treatment to be assigned each individual in a population. Then, treatment for individual j assigned treatment according to policy τ is denoted $z_{\tau j}$ and the corresponding outcome is $y_{\tau j} \equiv y\left(z_{\tau j}\right)$. As some individuals conform with their assignment and some may not, population outcomes under policy τ are described by random variable y_τ

$$y_\tau = \sum_t y\left(t\right) \cdot 1\left(z_\tau = t\right)$$

Conditional on covariates, x, the probability distribution for population outcomes is

$$\Pr\left(y_{\tau} \mid x\right) = \sum_{t} \Pr\left(y\left(t\right) \mid x, z_{\tau} = t\right) \Pr\left(z_{\tau} = t \mid x\right)$$

A classical experiment supplies an answer to the treatment effect question, while the mixing problem seeks the distribution of outcome under policy τ. A variety of policy implications are possible based on classical experimental evidence depending on what attendant conditions the analyst finds credible.

A randomized experiment reveals $\Pr\left(y\left(t\right) \mid x\right)$. The law of total probability says

$$\begin{aligned} \Pr\left(y\left(t\right) \mid x\right) &= \Pr\left(y\left(t\right) \mid x, z_{\tau} = t\right) \Pr\left(z_{\tau} = t \mid x\right) \\ &\quad + \Pr\left(y\left(t\right) \mid x, z_{\tau} \neq t\right) \Pr\left(z_{\tau} \neq t \mid x\right) \end{aligned}$$

Thus, $\Pr\left(y\left(t\right) \mid x\right)$ is a mixture of outcome distributions $\Pr\left(y\left(t\right) \mid x, z_{\tau} = t\right)$ and $\Pr\left(y\left(t\right) \mid x, z_{\tau} \neq t\right)$ with mixing probability weights $\Pr\left(z_{\tau} = t \mid x\right)$ and $\Pr\left(z_{\tau} \neq t \mid x\right)$. The mixing problem involves establishing bounds on the outcome distributions $\Pr\left(y\left(t\right) \mid x, z_{\tau} = t\right)$ for all treatments t from the marginal distributions $\Pr\left(y\left(t\right) \mid x\right)$ which, in turn, can be utilized to establish bounds on $\Pr\left(y_{\tau} \mid x\right)$. We consider both known and unknown mixing probabilities in a binary outcome, binary treatment example below. But first, we develop some preliminaries.

11.5.1 Fréchet bounds

Sometimes it's more convenient to work with joint outcome distributions than through outcome distributions conditional on treatment. This is the case, for instance, when questioning what classical experimental evidence alone tells the analyst about the mixing problem. In such cases, Fréchet bounds are instructive. Fréchet [1951] identified relations of marginal distributions that are helpful for bounding their joint distribution. Fréchet bounds are

$$\begin{aligned} &\max\left\{0, \Pr\left(y\left(t\right) \in A\right) + \Pr\left(y\left(s\right) \in A\right) - 1\right\} \\ \leq\ &\Pr\left(y\left(t\right) \in A \cap y\left(s\right) \in A\right) \leq \\ &\min\left\{\Pr\left(y\left(t\right) \in A\right), \Pr\left(y\left(s\right) \in A\right)\right\} \end{aligned}$$

The upper bound is straightforward as the intersection of sets cannot exceed the smaller of the two sets. The lower bound follows from the union of sets. That is,

$$\begin{aligned} 1 &\geq \Pr\left(y\left(t\right) \in A \cup y\left(s\right) \in A\right) \\ &= \Pr\left(y\left(t\right) \in A\right) + \Pr\left(y\left(s\right) \in A\right) - \Pr\left(y\left(t\right) \in A \cap y\left(s\right) \in A\right) \end{aligned}$$

Rearranging yields

$$\Pr(y(t) \in A \cap y(s) \in A) \geq \Pr(y(t) \in A) + \Pr(y(s) \in A) - 1$$

Since probability cannot be smaller than zero, we have Fréchet's lower bound.

11.5.2 Examples — binary outcome, binary treatment

Example 43 (binary outcome, binary treatment) *Suppose randomly assigned treatment is denoted* $t = t_1$ *and control or no treatment is denoted* $t = t_0$ *and outcomes are success* $y(t) = 1$ *or failure* $y(t) = 0$*. Further, suppose the evidence indicates* $\Pr(y(t_1) = 1 \mid x) = 0.67$ *and* $\Pr(y(t_0) = 1 \mid x) = 0.49$*. What does this classical experiment reveal about other treatment policies? Next, we explore various treatment response and treatment policy identification strategies for which the success probability identification regions are tabulated below.*

identification conditions	*success rate:* $\Pr(y_\tau = 1 \mid x)$
experimental evidence alone	$[0.16, 1]$
treatment response conditions:	
statistically independent outcomes	$[0.33, 0.83]$
monotone treatment response	$[0.49, 0.67]$
treatment policies:	
treatment at random	$[0.49, 0.67]$
outcome maximization	$[0.67, 1]$
outcome maximization and statistically independent outcomes	0.83
known mixing probabilities:	
$\frac{1}{10}$ *population receives treatment*	$[0.39, 0.59]$
$\frac{5}{10}$ *population receives treatment*	$[0.17, 0.99]$
$\frac{9}{10}$ *population receives treatment*	$[0.57, 0.77]$
$\frac{1}{10}$ *population receives treatment at random*	0.51
$\frac{5}{10}$ *population receives treatment at random*	0.58
$\frac{9}{10}$ *population receives treatment at random*	0.65

Clearly, a wide range of policy implications are possible depending on the conditions the analyst finds credible.

Example 44 (continuation of example 43 — evidence alone) *First, consider what the evidence alone tells the analyst about* $\Pr(y(t)=1 \mid x)$*. It may seem that if some individuals receive treatment and others do not that the success probability lies between* 0.49 *and* 0.67*. While this may be true under some conditions, it is not in general the case. Each individual faces one of four possibilities:* $[y(t_1)=0, y(t_0)=0]$, $[y(t_1)=1, y(t_0)=0]$, $[y(t_1)=0, y(t_0)=1]$*, and* $[y(t_1)=1, y(t_0)=1]$*. Varying treatment has no impact when* $y(t_1)=y(t_0)$ *but determines outcome when* $y(t_1)\neq y(t_0)$*. The highest success rate is achieved by a policy offering treatment* t_0 *to individuals with* $[y(t_1)=0, y(t_0)=1]$ *and treatment* t_1 *to individuals with* $[y(t_1)=1, y(t_0)=0]$*. Then, the only failure occurs for individuals with* $[y(t_1)=0, y(t_0)=0]$ *and success probability is* $1-\Pr(y(t_1)=0, y(t_0)=0)$*. Similarly, the lowest success rate is achieved by a policy offering treatment* t_1 *to individuals with* $[y(t_1)=0, y(t_0)=1]$ *and treatment* t_0 *to individuals with* $[y(t_1)=1, y(t_0)=0]$*. Then, the only individuals who achieve success are those with* $[y(t_1)=1, y(t_0)=1]$*. Hence, the lowest probability of success is* $\Pr(y(t_1)=1, y(t_0)=1)$*. This gives (suppressing the conditioning on covariates)*

$$\Pr(y(t_1)=1, y(t_0)=1) \leq \Pr(y_\tau=1) \leq 1-\Pr(y(t_1)=0, y(t_0)=0)$$

but the evidence doesn't directly reveal the joint probabilities as treatments are mutually exclusive. Rather, the experiment reveals marginal probabilities (again, omitting conditioning on covariates).

$$\Pr(y(t_1)=1)=0.67$$

and

$$\Pr(y(t_0)=1)=0.49$$

Using Bayes theorem, we search for the widest interval for $\Pr(y_\tau=1)$ *which involves assigning the set of conditional probabilities that minimize* $\Pr(y(t_1)=1, y(t_0)=1)+\Pr(y(t_1)=0, y(t_0)=0)$ *and are consistent with the marginal probabilities determined from the evidence. That is, we assign probabilities by solving the following constrained optimization problem. To simplify notation, let*

$$p_{ij} \equiv \Pr(y(t_0)=i \mid y(t_1)=j)$$

and

$$q_{ij} \equiv \Pr(y(t_1)=i \mid y(t_0)=j)$$

Now, solve

$$\begin{array}{cc} \min\limits_{p_{00},q_{00},q_{01},p_{01}\geq 0} & 0.51q_{00}+0.67(1-p_{01}) \\ & 0.51q_{00}-0.33p_{00}=0 \\ s.t. & 0.51(1-q_{00})-0.67p_{01}=0 \\ & 0.49q_{01}-0.33(1-p_{00})=0 \\ & 0.49(1-q_{01})-0.67(1-p_{01})=0 \end{array}$$

The solution is

$$\begin{aligned} p_{00} &\equiv \Pr\left(y\left(t_0\right)=0 \mid y\left(t_1\right)=0\right)=0 \\ q_{00} &\equiv \Pr\left(y\left(t_1\right)=0 \mid y\left(t_0\right)=0\right)=0 \\ q_{01} &\equiv \Pr\left(y\left(t_1\right)=0 \mid y\left(t_0\right)=1\right)=0.6735 \\ p_{01} &\equiv \Pr\left(y\left(t_0\right)=0 \mid y\left(t_1\right)=1\right)=0.7612 \end{aligned}$$

and implies the following joint outcomes distribution

$$\begin{aligned} \Pr\left(y\left(t_1\right)=0, y\left(t_0\right)=0\right) &= 0 \\ \Pr\left(y\left(t_1\right)=1, y\left(t_0\right)=0\right) &= 0.51 \\ \Pr\left(y\left(t_1\right)=0, y\left(t_0\right)=1\right) &= 0.33 \\ \Pr\left(y\left(t_1\right)=1, y\left(t_0\right)=1\right) &= 0.16 \end{aligned}$$

Hence, the identification bounds from the evidence alone is

$$\begin{aligned} 0.16 &= \Pr\left(y\left(t_1\right)=1 \cap y\left(t_0\right)=1\right) \\ &\leq \Pr\left(y_\tau=1\right) \leq \\ &\quad 1-\Pr\left(y\left(t_1\right)=0 \cap y\left(t_0\right)=0\right)=1 \end{aligned}$$

This is an application of the Fréchet bound. The Fréchet bounds on the joint outcomes are

$$0.16 \leq \Pr\left(y\left(t_1\right)=1 \cap y\left(t_0\right)=1\right) \leq 0.49$$

and

$$0 \leq \Pr\left(y\left(t_1\right)=0 \cap y\left(t_0\right)=0\right) \leq 0.33$$

Hence, the evidence only slightly narrows the mixing bounds as the widest interval from the above Fréchet bounds is

$$0.16 \leq \Pr\left(y_\tau=1\right) \leq 1-0=1$$

Now, we explore conditions regarding response to treatment.

Example 45 (continuation of example 43 — independent outcomes) *Suppose outcomes are believed to be independent, then the identification region based on the data can be narrowed somewhat. The analyst appends the independence condition*

$$\begin{aligned} \Pr\left(y\left(t_0\right) \mid y\left(t_1\right)=0\right) &= \Pr\left(y\left(t_0\right) \mid y\left(t_1\right)=1\right) \\ \Pr\left(y\left(t_1\right) \mid y\left(t_0\right)=0\right) &= \Pr\left(y\left(t_1\right) \mid y\left(t_0\right)=1\right) \end{aligned}$$

or

$$\Pr\left(y\left(t_1\right), y\left(t_0\right)\right)=\Pr\left(y\left(t_1\right)\right)\Pr\left(y\left(t_0\right)\right)$$

to the foregoing analysis and again solve for the widest interval

$$\Pr\left(y\left(t_1\right)=1,y\left(t_0\right)=1\right)\leq\Pr\left(y_\tau=1\right)\leq1-\Pr\left(y\left(t_1\right)=0,y\left(t_0\right)=0\right)$$

Using Bayes theorem, this interval is widest for the following conditional distributions

$$\begin{aligned}\Pr\left(y\left(t_0\right)=0\mid y\left(t_1\right)=0\right) &= \Pr\left(y\left(t_0\right)=0\mid y\left(t_1\right)=1\right)=0.51\\ \Pr\left(y\left(t_1\right)=0\mid y\left(t_0\right)=0\right) &= \Pr\left(y\left(t_1\right)=0\mid y\left(t_0\right)=1\right)=0.33\end{aligned}$$

which implies the following joint distribution

$$\begin{aligned}\Pr\left(y\left(t_1\right)=0,y\left(t_0\right)=0\right) &= 0.17\\ \Pr\left(y\left(t_1\right)=1,y\left(t_0\right)=0\right) &= 0.34\\ \Pr\left(y\left(t_1\right)=0,y\left(t_0\right)=1\right) &= 0.16\\ \Pr\left(y\left(t_1\right)=1,y\left(t_0\right)=1\right) &= 0.33\end{aligned}$$

Hence, outcome independence in conjunction with the evidence yields $0.33\leq\Pr\left(y_\tau=1\right)\leq0.83$*.*

Example 46 (cont. of example 43 — monotone treatment response) *If the analyst believes treatment is never harmful to any individual, then treatment response is monotonic,* $y\left(t_1\right)\geq y\left(t_0\right)$ *for all individuals. This is an MTR identification strategy and success probability lies between that observed for the untreated and that observed for the treated, Namely,* $0.49\leq\Pr\left(y_\tau=1\right)\leq0.67$*. To see this recall MTR implies* $\Pr\left(y\left(t_1\right)\right)$ *stochastically dominates* $\Pr\left(y\left(t_0\right)\right)$*. That is,*

$$\Pr\left(y\left(t_1\right)\leq0\right)\leq\Pr\left(y_\tau\leq0\right)\leq\Pr\left(y\left(t_0\right)\leq0\right)$$

can be translated as $\Pr\left(y_\tau=1\right)=1-\Pr\left(y_\tau\leq0\right)$ *resulting in*

$$0.49\leq\Pr\left(y_\tau=1\right)\leq0.67$$

Notice, evidence from randomized treatment may refute MTR as the interval may be empty.

Next, we turn attention from treatment response to different treatment policies.

Example 47 (cont. of example 43 — treatment at random) *If the analyst wishes to explore a policy of randomized treatment then treatment is independent of outcome* $\left[y\left(t_1\right),y\left(t_0\right)\right]$*. In other words,*

$$\Pr\left(y\left(t_1\right),y\left(t_0\right)\mid z_\tau\right)=\Pr\left(y\left(t_1\right),y\left(t_0\right)\right)$$

This implies

$$\Pr\left(y_\tau\right)=\Pr\left(y\left(t_1\right)\mid z_\tau=t_1\right)\Pr\left(z_\tau=t_1\right)+\Pr\left(y\left(t_0\right)\mid z_\tau=t_0\right)\Pr\left(z_\tau=t_0\right)$$

which simplifies as

$$\Pr(y_\tau) = \Pr(y(t_1))\Pr(z_\tau = t_1) + \Pr(y(t_0))\Pr(z_\tau = t_0)$$

If $\Pr(z_\tau)$ *is known then* $\Pr(y_\tau)$ *is point-identified. Otherwise, the bounds are*

$$\begin{array}{rl} & \min\{\Pr(y(t_1) = 1), \Pr(y(t_0) = 1)\} \\ \leq & \Pr(y_\tau) \leq \\ & \max\{\Pr(y(t_1) = 1), \Pr(y(t_0) = 1)\} \end{array}$$

or

$$0.49 \leq \Pr(y_\tau) \leq 0.67$$

Example 48 (cont. of example 43 — outcome optimized) *Suppose the policy is to treat according to best response*

$$y_\tau = \max\{y(t_1), y(t_0)\}$$

This implies

$$\Pr(y_\tau \leq 0) = \Pr(y(t_1) \leq 0 \cap y(t_0) \leq 0)$$

The Fréchet bound indicates the right hand side involves bounds

$$\begin{array}{rl} & \max\{0, \Pr(y(t_1) \leq 1) + \Pr y((t_0) \leq 1) - 1\} \\ \leq & \Pr(y(t_1) \leq 0 \cap y(t_0) \leq 0) \leq \\ & \min\{\Pr(y(t_1) \leq 1), \Pr(y((t_0) \leq 1))\} \end{array}$$

Again, applying the translation $\Pr(y_\tau = 1) = 1 - \Pr(y_\tau \leq 0)$ *we have*

$$\begin{array}{rl} & 1 - \min\{\Pr(y(t_1) \leq 1), \Pr(y(t_0) \leq 1)\} \\ \leq & 1 - \Pr(y_\tau \leq 0) \leq \\ & 1 - \max\{0, \Pr(y(t_1) \leq 1) + \Pr y((t_0) \leq 1) - 1\} \end{array}$$

$$\begin{array}{rcl} 1 - 0.33 & \leq & \Pr(y_\tau = 1) \leq 1 - \max\{0, 0.84 - 1\} \\ 0.67 & \leq & \Pr(y_\tau = 1) \leq 1 \end{array}$$

Example 49 (cont. of example 43 — optimal independent outcomes) *Suppose the policy treats according to best response and outcomes are independent. Then,*

$$\begin{array}{rcl} \Pr(y_\tau \leq 0) & = & \Pr(y(t_1) \leq 1)\Pr(y(t_0) \leq 1) \\ & = & 0.33 \cdot 0.51 = 0.17 \end{array}$$

and

$$\Pr(y_\tau = 1) = 0.83$$

In other words, the probability is point-identified.

Next, we explore some cases involving known mixing probabilities.

Example 50 (cont. of example 43 — known percent treated) *Suppose a known fraction p of the population receives treatment* $\Pr(z_\tau = t_1) = p$. *The law of total probability gives*

$$\begin{aligned}\Pr(y_\tau = 1) &= \Pr(y(t_1) = 1 \mid z_\tau = t_1)\, p \\ &\quad + \Pr(y(t_0) = 1 \mid z_\tau = t_0)(1 - p)\end{aligned}$$

The experimental evidence, $\Pr(y(t_0) = 1)$ *and* $\Pr(y(t_1) = 1)$, *relates to the quantities of interest,* $\Pr(y(t_0) = 1 \mid z_\tau = t_0)$ *and* $\Pr(y(t1) = 1 \mid z_\tau = t_0)$, *as follows.*

$$\begin{aligned}\Pr(y(t_0) = 1) &= \Pr(y(t_0) = 1 \mid z_\tau = t_0)(1 - p) \\ &\quad + \Pr(y(t_0) = 1 \mid z_\tau = t_1)\, p\end{aligned}$$

and

$$\begin{aligned}\Pr(y(t_1) = 1) &= \Pr(y(t_1) = 1 \mid z_\tau = t_0)(1 - p) \\ &\quad + \Pr(y(t_1) = 1 \mid z_\tau = t_1)\, p\end{aligned}$$

Let $C_0 \equiv \Pr(y(t_0) = 1 \mid z_\tau = t_1)$ *and* $C_1 \equiv \Pr(y(t_1) = 1 \mid z_\tau = t_0)$. *If* $C_0 = 1$, *then the lower bound is*

$$\Pr(y(t_0) = 1 \mid z_\tau = t_0) = \max\left\{0, \frac{\Pr(y(t_0) = 1) - p}{1 - p}\right\}$$

If $C_0 = 0$, *then the upper bound is*

$$\Pr(y(t_0) = 1 \mid z_\tau = t_0) = \min\left\{1, \frac{\Pr(y(t_0) = 1)}{1 - p}\right\}$$

Similarly, if $C_1 = 1$, *then the lower bound is*

$$\Pr(y(t_1) = 1 \mid z_\tau = t_1) = \max\left\{0, \frac{\Pr(y(t_1) = 1) - (1 - p)}{p}\right\}$$

If $C_1 = 0$, *then the upper bound is*

$$\Pr(y(t_1) = 1 \mid z_\tau = t_1) = \min\left\{1, \frac{\Pr(y(t_1) = 1)}{p}\right\}$$

Thus, the identification region for

$$\begin{aligned}\Pr(y_\tau = 1) &= \Pr(y(t_1) = 1 \mid z_\tau = t_1)\, p \\ &\quad + \Pr(y(t_0) = 1 \mid z_\tau = t_0)(1 - p)\end{aligned}$$

is

$$\begin{aligned} & \max\{0, \Pr(y(t_0)=1)-p\} + \max\{0, \Pr(y(t_1)=1)-(1-p)\} \\ \leq \; & \Pr(y_\tau = 1) \leq \\ & \min\{1-p, \Pr(y(t_0)=1)\} + \min\{p, \Pr(y(t_1)=1)\} \end{aligned}$$

Therefore, for $p = 0.10$*, we have*

$$\begin{aligned} & & & \max\{0, 0.49-0.10\} + \max\{0, 0.67-0.90\} \\ & \leq & & \Pr(y_\tau = 1) \leq \\ & & & \min\{0.90, 0.49\} + \min\{0.10, 0.67\} \\ 0.39 & \leq & & \Pr(y_\tau = 1) \leq 0.59 \end{aligned}$$

For $p = 0.50$*, we have*

$$\begin{aligned} & & & \max\{0, 0.49-0.50\} + \max\{0, 0.67-0.50\} \\ & \leq & & \Pr(y_\tau = 1) \leq \\ & & & \min\{0.50, 0.49\} + \min\{0.50, 0.67\} \\ 0.17 & \leq & & \Pr(y_\tau = 1) \leq 0.99 \end{aligned}$$

For $p = 0.90$*, we have*

$$\begin{aligned} & & & \max\{0, 0.49-0.90\} + \max\{0, 0.67-0.10\} \\ & \leq & & \Pr(y_\tau = 1) \leq \\ & & & \min\{0.10, 0.49\} + \min\{0.90, 0.67\} \\ 0.57 & \leq & & \Pr(y_\tau = 1) \leq 0.77 \end{aligned}$$

Example 51 (cont. of example 43 — randomly treat known percent) *Suppose the policy is to treat fraction* p *of the population at random. Then,*

$$\Pr(y_\tau = 1) = \Pr(y(t_0))(1-p) + \Pr(y(t_1))\,p$$

Since all of the right hand side quantities are known, $\Pr(y_\tau = 1)$ *is point identified. For* $p = 0.10$*, we have*

$$\Pr(y_\tau = 1) = 0.49 \cdot 0.90 + 0.67 \cdot 0.10 = 0.51$$

For $p = 0.50$*, we have*

$$\Pr(y_\tau = 1) = 0.49 \cdot 0.50 + 0.67 \cdot 0.50 = 0.58$$

For $p = 0.90$*, we have*

$$\Pr(y_\tau = 1) = 0.49 \cdot 0.10 + 0.67 \cdot 0.90 = 0.65$$

11.6 Appendix: bounds on spreads

We briefly discuss nonparametric identification bounds for spread parameters following Blundell, Gosling, Ichimura, and Meghir [*Econometrica*, 2007] (BGIM). Suppose the analyst is concerned about the spread in the distribution of some random variable y (the classic case, spread in wages, is addressed by BGIM but we can imagine reported quantities, like spread in earnings, to be the object of interest as well). Nonparametric spreads are derived from differences in quantiles such as the interquartile range. Let the quantiles be denoted by q_i where $q_2 > q_1$. The conservative identification region for the spread, $D = y^{q_2} - y^{q_1}$,[16] is

$$\max\left\{0, y^{q_2(l)} - y^{q_1(u)}\right\} \leq D \leq y^{q_2(u)} - y^{q_1(l)}$$

where $y^{q_i(l)}$ is the lower bound for quantile q_i of y, and $y^{q_i(u)}$ is the upper bound for quantile q_i of y. The bounds for quantiles

$$y^{q(l)} \leq y^q \leq y^{q(u)}$$

are derived from bounds on the *CDF* (cumulative distribution function)

$$\Pr(z=1)\, F(y \mid z=1) \leq F(y) \leq \Pr(z=1)\, F(y \mid z=1) + \Pr(z=0)$$

In other words, the lower bound solves

$$q = \Pr(z=1)\, F(y \mid z=1) + \Pr(z=0)$$

for y, while the upper bound solves

$$q = \Pr(z=1)\, F(y \mid z=1)$$

for y. Without restrictions on the support of Y, the analyst can only identify lower bounds for $q > \Pr(z=0)$ and upper bounds for $q < \Pr(z=1)$.

BGIM exploit properties of the *CDF* to narrow bounds on D. By the law of total probability

$$F(y) = \Pr(z=1)\, F(y \mid z=1) + \Pr(z=0)\, F(y \mid z=0)$$

which can be rewritten as

$$F(y \mid z=0) = \frac{F(y) - \Pr(z=1)\, F(y \mid z=1)}{\Pr(z=0)}$$

Recognizing the *CDF* for missing data, $F(y \mid z=0)$, is nondecreasing in y implies $F(y) \geq \Pr(z=1)\, F(y \mid z=1)$. Now, express quantile q_2 in terms

[16] Throughout this discussion, conditioning on covariates is suppressed to simplify notation.

of the distribution for the observable data relative to some uncertain quantile q_1 value $y_0 \in \left[y^{q_1(l)}, y^{q_1(u)}\right]$

$$q_2 = \Pr(z=1)\left[F(y \mid z=1) - F(y_0 \mid z=1)\right] + q_1$$

where the quantity in brackets is nonnegative. Rearranging and utilizing Bayes rule gives

$$F(y \mid z=1) = F(y_0 \mid z=1) + \frac{q_2 - q_1}{\Pr(z=1)}$$

This is the upper bound on the q_2-quantile relative to q_1 expressed in terms of the uncertain q_1-quantile value, y_0. Then, the upper bound on the spread is (weakly) less than the conservative upper bound

$$\begin{aligned} D^{(u)} &= \sup_{y_0 \in \left[y^{q_1(l)}, y^{q_1(u)}\right]} \left\{ F^{-1}\left(F(y_0 \mid z=1) + \frac{q_2 - q_1}{\Pr(z=1)} \right) - y_0 \right\} \\ &\leq y^{q_2(u)} - y^{q_1(l)} \end{aligned}$$

The lower bound for the spread is simply

$$D^{(l)} = \max\left\{0, y^{q_2(l)} - y^{q_1(u)}\right\}$$

the same as the conservative lower bound. Thus, the BGIM *CDF*-based identification bounds for spread

$$D^{(l)} \leq D \leq D^{(u)}$$

are (weakly) tighter than the conservative bounds

$$\max\left\{0, y^{q_2(l)} - y^{q_1(u)}\right\} \leq D \leq y^{q_2(u)} - y^{q_1(l)}$$

Next, we'll illustrate identification bounds for spread parameters (specifically, the interquartile range, $D = y^{75} - y^{25}$) via some examples. In order to demonstrate the bounds, we consider examples in which the missing outcome data lie entirely above or entirely below the observable outcome data.

11.6.1 Examples

Example 52 $\left((y \mid z=0) < (y \mid z=1)\right)$ *Suppose outcome is normally distributed with mean* $1,000$*, standard deviation* 300*, and interquartile range* $1202 - 798 = 405$ *where the analyst observes the top* p *fraction of the population. In other words, the missing data are all from the bottom* $(1-p)$

fraction and the observable distribution is truncated below. The following table depicts the BGIM and conservative bounds for various p.

	conservative bounds		*BGIM bounds*	
p	$D^{(l)}$	$D^{(u)}$	$D^{(l)}$	$D^{(u)}$
0.9	318	513	318	427
0.8	240	696	240	531
0.76	210	900	210	705
0.7	165	*unbounded*	165	*unbounded*
0.6	87	*unbounded*	87	*unbounded*
0.5	0	*unbounded*	0	*unbounded*
0.4	0	*unbounded*	0	*unbounded*

p	$y^{q_{25}(l)}$	$y^{q_{25}(u)}$	$y^{q_{75}(l)}$	$y^{q_{75}(u)}$
0.9	798	884	1202	1311
0.8	798	962	1202	1493
0.76	798	992	1202	1698
0.7	*unbounded*	1038	1202	*unbounded*
0.6	*unbounded*	1116	1202	*unbounded*
0.5	*unbounded*	1202	1202	*unbounded*
0.4	*unbounded*	1311	1202	*unbounded*

Naturally, the bounds are inversely related to p as the information or evidence available to the analyst is increasing in p.

Example 53 $((y \mid z = 0) > (y \mid z = 1))$ *Suppose the DGP remains as above except the truncation occurs above rather than below. In other words, the top $(1 - p)$ fraction of outcomes are missing. The table below describes the BGIM and conservative bounds for various p.*

	conservative bounds		*BGIM bounds*	
p	$D^{(l)}$	$D^{(u)}$	$D^{(l)}$	$D^{(u)}$
0.9	318	513	318	427
0.8	240	696	240	531
0.76	210	900	210	705
0.7	165	*unbounded*	165	*unbounded*
0.6	87	*unbounded*	87	*unbounded*
0.5	0	*unbounded*	0	*unbounded*
0.4	0	*unbounded*	0	*unbounded*

p	$y^{q_{25}(l)}$	$y^{q_{25}(u)}$	$y^{q_{75}(l)}$	$y^{q_{75}(u)}$
0.9	689	798	1116	1202
0.8	507	798	1038	1202
0.76	302	798	1008	1202
0.7	*unbounded*	798	962	*unbounded*
0.6	*unbounded*	798	884	*unbounded*
0.5	*unbounded*	798	798	*unbounded*
0.4	*unbounded*	798	689	*unbounded*

With symmetrical lower and upper truncation in the two examples, the analyst observes a shift from the lower quantile-25 *to upper quantile*-25 *bound of* 798 *and a similar shift from lower quantile*-75 *bound to upper quantile*-75 *bound of* 1202. *The interquartile range is unbounded when half or more of the data are missing at either end of the distribution.*