Notes for Paper on Mis-measured, Binary, Endogenous Regressors

Francis J. DiTraglia & Camilo García-Jimeno

June 3, 2017

1 Model and Notation

Probabilities

$$
\begin{aligned}
p^*_{tk} &= P(T^* = t, Z = k) \\
p_{tk} &= P(T = t, Z = k) \\
p^*_k &= P(T^* = 1|Z = k) \\
p_k &= P(T = 1|Z = k) \\
q &= P(Z = 1)
\end{aligned}
$$

$$
\begin{aligned}
p^*_{00} &= P(T^* = 0|Z = 0)P(Z = 0) = (1 - p^*_0)(1 - q) = \left(\frac{1 - p_0 - \alpha_1}{1 - \alpha_0 - \alpha_1}\right)(1 - q) \\
p^*_{10} &= P(T^* = 1|Z = 0)P(Z = 0) = p^*_0(1 - q) = \left(\frac{p_0 - \alpha_0}{1 - \alpha_0 - \alpha_1}\right)(1 - q) \\
p^*_{01} &= P(T^* = 0|Z = 1)P(Z = 1) = (1 - p^*_1)q = \left(\frac{1 - p_1 - \alpha_1}{1 - \alpha_0 - \alpha_1}\right)q \\
p^*_{11} &= P(T^* = 1|Z = 1)P(Z = 1) = p^*_1(1 - q) = \left(\frac{p_1 - \alpha_0}{1 - \alpha_0 - \alpha_1}\right)q
\end{aligned}
$$

CDFs For $t, Z \in \{0, 1\}$ define

$$
\begin{aligned}
F^*_{tk}(\tau) &= P(Y \leq \tau|T^* = t, Z = k) \\
F_{tk}(\tau) &= P(Y \leq \tau|T = t, Z = k) \\
F_k(\tau) &= P(Y \leq \tau|Z = k)
\end{aligned}
$$

Note that the second two are observed for all t, k while the first is never observed since it depends on the unobserved RV T^*.

2 Weakest Bounds on α_0, α_1

Assume that $\alpha_0 + \alpha_1 < 1$ that T is independent of Z conditional on T^*. These standard assumptions turn out to yield informative bounds on α_0 and α_1 without *any further restrictions of any kind*. In particular, we assume nothing about the validity of the instrument Z and nothing about the relationship between the mis-classification error and the outcome Y: we impose only that the mis-classification error rates do not depend on z and that the mis-classification is not so bad that $1 - T$ is a better measure of T^* than T.

By the Law of Total Probability and the assumption that T is conditionally independent of Z given T^*,

$$
\begin{aligned}
p_k &= P(T = 1|Z = k, T^* = 0)(1 - p_k^*) + P(T = 1|Z = k, T^* = 1)p_k^* \\
&= P(T = 1|T^* = 0)(1 - p_k^*) + P(T = 1|T^* = 1)p_k^* \\
&= \alpha_0(1 - p_k^*) + (1 - \alpha_1)p_k^* \\
&= \alpha_0 + (1 - \alpha_0 - \alpha_1)p_k^*
\end{aligned}
$$

and similarly

$$
\begin{aligned}
1 - p_k &= P(T = 0|Z = k, T^* = 0)(1 - p_k^*) + P(T = 0|Z = k, T^* = 1)p_k^* \\
&= P(T = 0|T^* = 0)(1 - p_k^*) + P(T = 0|T^* = 1)p_k^* \\
&= (1 - \alpha_0)(1 - p_k^*) + \alpha_1 p_k^* \\
&= \alpha_1 + (1 - p_k^*)(1 - \alpha_0 - \alpha_1)
\end{aligned}
$$

and hence

$$
\begin{aligned}
p_k - \alpha_0 &= (1 - \alpha_0 - \alpha_1)p_k^* \\
(1 - p_k) - \alpha_1 &= (1 - \alpha_0 - \alpha_1)(1 - p_k^*)
\end{aligned}
$$

Now, since p_k^* and $(1 - p_k^*)$ are probabilities they are between zero and one which means that the sign of $p_k - \alpha_0$ as well as that of $(1 - p_k) - \alpha_1$ are both determined by that of $1 - \alpha_0 - \alpha_1$.

Accordingly, provided that $1 - \alpha_0 - \alpha_1 < 1$, we have

$$\begin{aligned} \alpha_0 &< p_k \\ \alpha_1 &< (1 - p_k) \end{aligned}$$

so long as p_k^* does not equal zero or one, which is not a realistic case for any example that we consider. Since these bounds hold for all k, we can take the tightest bound over all values of Z.

Important: using these to bound β gives $\beta \in [\text{ITT}, \text{Wald}]$.

3 Stronger Bounds for α_0, α_1

Now suppose we add the assumption that T is conditionally independent of Y given T^*. This is essentially the non-differential measurement error assumption although it is slightly stronger than the version used by Mahajan (2006) who assumes only conditional mean independence. This assumption allows us to considerably strengthen the bounds from the preceding section by exploiting information contained in the conditional distribution of Y given T and Z. The key ingredient is a relationship that we can derive between the unobservable distributions F_{tk}^* and the observable distributions F_{tk} using this new conditional independence assumption. To begin, note that by Bayes' rule we have

$$\begin{aligned} P(T^* = 1 | T = 1, Z = k) &= P(T = 1 | T^* = 1) \left(\frac{p_k^*}{p_k} \right) = (1 - \alpha_1) \left(\frac{p_k^*}{p_k} \right) \\ P(T^* = 1 | T = 0, Z = k) &= P(T = 0 | T^* = 1) \left(\frac{p_k^*}{1 - p_k} \right) = \alpha_1 \left(\frac{p_k^*}{1 - p_k} \right) \\ P(T^* = 0 | T = 1, Z = k) &= P(T = 1 | T^* = 0) \left(\frac{1 - p_k^*}{p_k} \right) = \alpha_0 \left(\frac{1 - p_k^*}{p_k} \right) \\ P(T^* = 0 | T = 0, Z = k) &= P(T = 0 | T^* = 0) \left(\frac{1 - p_k^*}{1 - p_k} \right) = (1 - \alpha_0) \left(\frac{1 - p_k^*}{1 - p_k} \right) \end{aligned}$$

Now, by the conditional independence assumption

$$\begin{aligned} P(Y \leq \tau | T^* = 0, T = t, Z = k) = P(Y \leq \tau | T^* = 0, Z = k) = F_{0k}^*(\tau) \\ P(Y \leq \tau | T^* = 1, T = t, Z = k) = P(Y \leq \tau | T^* = 1, Z = k) = F_{1k}^*(\tau) \end{aligned}$$

Finally, putting everything together using the Law of Total Probability, we find that

$$\begin{aligned}(1-p_k)F_{0k}(\tau) &= (1-\alpha_0)(1-p_k^*)F_{0k}^*(\tau) + \alpha_1 p_k^* F_{1k}^*(\tau)\\ p_k F_{1k}(\tau) &= \alpha_0(1-p_k^*)F_{0k}^*(\tau) + (1-\alpha_1)p_k^* F_{1k}^*(\tau)\end{aligned}$$

for all k. Defining the shorthand

$$\begin{aligned}\widetilde{F}_{0k}(\tau) &\equiv (1-p_k)F_{0k}(\tau)\\ \widetilde{F}_{1k}(\tau) &\equiv p_k F_{1k}(\tau)\end{aligned}$$

this becomes

$$\widetilde{F}_{0k}(\tau) = (1-\alpha_0)(1-p_k^*)F_{0k}^*(\tau) + \alpha_1 p_k^* F_{1k}^*(\tau) \tag{1}$$

$$\widetilde{F}_{1k}(\tau) = \alpha_0(1-p_k^*)F_{0k}^*(\tau) + (1-\alpha_1)p_k^* F_{1k}^*(\tau) \tag{2}$$

Now, solving Equation 1 for $p_k^* F_{1k}^*(\tau)$ we have

$$p_k^* F_{1k}^*(\tau) = \frac{1}{\alpha_1}\left[\widetilde{F}_{0k}(\tau) - (1-\alpha_0)(1-p_k^*)F_{0k}^*(\tau)\right]$$

Substituting this into Equation 2,

$$\begin{aligned}\widetilde{F}_{1k}(\tau) &= \alpha_0(1-p_k^*)F_{0k}^*(\tau) + \frac{1-\alpha_1}{\alpha_1}\left[\widetilde{F}_{0k}(\tau) - (1-\alpha_0)(1-p_k^*)F_{0k}^*(\tau)\right]\\ &= \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) + \left[\alpha_0 - \frac{(1-\alpha_1)(1-\alpha_0)}{\alpha_1}\right](1-p_k^*)F_{0k}^*(\tau)\\ &= \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) + \left[\frac{\alpha_0\alpha_1 - (1-\alpha_1)(1-\alpha_0)}{\alpha_1}\right](1-p_k^*)F_{0k}^*(\tau)\\ &= \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) - \left[\frac{(1-\alpha_1)(1-\alpha_0) - \alpha_0\alpha_1}{\alpha_1}\right](1-p_k^*)F_{0k}^*(\tau)\\ &= \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) - \left[\frac{1-\alpha_1-\alpha_0}{\alpha_1}\right]\left(\frac{1-p_k-\alpha_1}{1-\alpha_0-\alpha_1}\right)F_{0k}^*(\tau)\end{aligned}$$

and therefore

$$\widetilde{F}_{1k}(\tau) = \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) - \frac{1-p_k-\alpha_1}{\alpha_1}F_{0k}^*(\tau) \tag{3}$$

Equation 3 relates the observable $\widetilde{F}_{1k}(\tau)$ to the mis-classification error rate α_1 and the unobservable CDF $F_{0k}^*(\tau)$. Since $F_{0k}^*(\tau)$ is a CDF, however, it lies in the interval $[0,1]$.

Accordingly, substituting 0 in place of $F_{0k}^*(\tau)$ gives

$$\widetilde{F}_{1k}(\tau) \leq \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) \tag{4}$$

while substituting 1 gives

$$\widetilde{F}_{1k}(\tau) \geq \frac{1-\alpha_1}{\alpha_1}\widetilde{F}_{0k}(\tau) - \frac{1-p_k-\alpha_1}{\alpha_1} \tag{5}$$

Rearranging Equation 4

$$\begin{aligned}
\alpha_1 \widetilde{F}_{1k}(\tau) &\leq (1-\alpha_1)\widetilde{F}_{0k}(\tau) \\
\alpha_1 \widetilde{F}_{1k}(\tau) &\leq \widetilde{F}_{0k}(\tau) - \alpha_1 \widetilde{F}_{0k}(\tau) \\
\alpha_1 \left[\widetilde{F}_{0k}(\tau) + \widetilde{F}_{1k}(\tau)\right] &\leq \widetilde{F}_{0k}(\tau)
\end{aligned}$$

since $\alpha_1 \in [0,1]$ and therefore

$$\alpha_1 \leq \frac{\widetilde{F}_{0k}(\tau)}{\widetilde{F}_{0k}(\tau) + \widetilde{F}_{1k}(\tau)} = (1-p_k)\left[\frac{F_{0k}(\tau)}{F_k(\tau)}\right] \tag{6}$$

since $\widetilde{F}_{1k}(\tau) + \widetilde{F}_{1k}(\tau) \geq 0$. Proceeding similarly for Equation 5,

$$\begin{aligned}
\alpha_1 \widetilde{F}_{1k}(\tau) &\geq (1-\alpha_1)\widetilde{F}_{0k}(\tau) - (1-p_k-\alpha_1) \\
\alpha_1 \left[\widetilde{F}_{1k}(\tau) + \widetilde{F}_{0k}(\tau) - 1\right] &\geq \widetilde{F}_{0k}(\tau) - (1-p_k) \\
-\alpha_1 \left[1 - \widetilde{F}_{1k}(\tau) - \widetilde{F}_{0k}(\tau)\right] &\geq -\left[1 - \widetilde{F}_{0k}(\tau) - p_k\right] \\
\alpha_1 \left[1 - \widetilde{F}_{1k}(\tau) - \widetilde{F}_{0k}(\tau)\right] &\leq 1 - \widetilde{F}_{0k}(\tau) - p_k
\end{aligned}$$

Now since $\widetilde{F}_{1k}(\tau) = p_k F_{1k}(\tau) \leq p_k$ and $\widetilde{F}_{0k}(\tau) = (1-p_k)F_{0k}(\tau) \leq (1-p_k)$ it follows that $1 - \widetilde{F}_{1k}(\tau) - \widetilde{F}_{0k}(\tau) \geq 0$ and hence

$$\alpha_1 \leq \frac{1 - \widetilde{F}_{0k}(\tau) - p_k}{1 - \widetilde{F}_{1k}(\tau) - \widetilde{F}_{0k}(\tau)} = (1-p_k)\left[\frac{1 - F_{0k}(\tau)}{1 - F_k(\tau)}\right] \tag{7}$$

The bounds given in Equations 6 and 7 relate α_1 to observable quantities *only* and hold for all values of τ for which their respective denominators are non-zero. Moreover, these bounds hold for any value k that the instrument takes on.

We can proceed similarly for α_0. First solve Equation 1 for $(1-p_k^*)F_{0k}^*(\tau)$:

$$(1-p_k^*)F_{0k}^*(\tau) = \frac{1}{1-\alpha_0}\left[\widetilde{F}_{0k}(\tau) - \alpha_1 p_k^* F_{1k}^*(\tau)\right]$$

and then substitute into Equation 2:

$$\begin{aligned}
\widetilde{F}_{1k}(\tau) &= \frac{\alpha_0}{1-\alpha_0}\left[\widetilde{F}_{0k}(\tau) - \alpha_1 p_k^* F_{1k}^*(\tau)\right] + (1-\alpha_1)p_k^* F_{1k}^*(\tau) \\
&= \frac{\alpha_0}{1-\alpha_0}\widetilde{F}_{0k}(\tau) + \left[(1-\alpha_1) - \frac{\alpha_0\alpha_1}{1-\alpha_0}\right] p_k^* F_{1k}^*(\tau) \\
&= \frac{\alpha_0}{1-\alpha_0}\widetilde{F}_{0k}(\tau) + \left[\frac{(1-\alpha_1)(1-\alpha_0) - \alpha_0\alpha_1}{1-\alpha_0}\right] p_k^* F_{1k}^*(\tau) \\
&= \frac{\alpha_0}{1-\alpha_0}\widetilde{F}_{0k}(\tau) + \left[\frac{1-\alpha_0-\alpha_1}{1-\alpha_0}\right] \frac{p_k-\alpha_0}{1-\alpha_0-\alpha_1} F_{1k}^*(\tau)
\end{aligned}$$

and therefore

$$\widetilde{F}_{1k}(\tau) = \frac{\alpha_0}{1-\alpha_0}\widetilde{F}_{0k}(\tau) + \frac{p_k-\alpha_0}{1-\alpha_0} F_{1k}^*(\tau) \tag{8}$$

Now we can again obtain two bounds by substituting the smallest and largest possible values of $F_{1k}^*(\tau)$. Substituting zero gives

$$\widetilde{F}_{1k}(\tau) \geq \frac{\alpha_0}{1-\alpha_0}\widetilde{F}_{0k}(\tau) \tag{9}$$

while substituting one gives

$$\widetilde{F}_{1k}(\tau) \leq \frac{\alpha_0}{1-\alpha_0}\widetilde{F}_{0k}(\tau) + \frac{p_k-\alpha_0}{1-\alpha_0} \tag{10}$$

Now, rearranging Equation 9,

$$\begin{aligned}
(1-\alpha_0)\widetilde{F}_{1k}(\tau) &\geq \alpha_0\widetilde{F}_{0k}(\tau) \\
\widetilde{F}_{1k}(\tau) &\geq \alpha_0\left[\widetilde{F}_{0k}(\tau) + \widetilde{F}_{1k}(\tau)\right]
\end{aligned}$$

since $1-\alpha_0 \geq 0$. Therefore,

$$\alpha_0 \leq \frac{\widetilde{F}_{1k}(\tau)}{\widetilde{F}_{0k}(\tau) + \widetilde{F}_{1k}(\tau)} = p_k\left[\frac{F_{1k}(\tau)}{F_k(\tau)}\right] \tag{11}$$

since $\left[\widetilde{F}_{0k}(\tau)+\widetilde{F}_{1k}(\tau)\right] \geq 0$. Similarly, rearranging Equation 10

$$
\begin{aligned}
(1-\alpha_0)\widetilde{F}_{1k}(\tau) &\leq \alpha_0 \widetilde{F}_{0k}(\tau) + p_k - \alpha_0 \\
\widetilde{F}_{1k}(\tau) - p_k &\leq \alpha_0\left[\widetilde{F}_{0k}(\tau)+\widetilde{F}_{1k}(\tau) - 1\right] \\
-\left[1-\widetilde{F}_{1k}(\tau)-(1-p_k)\right] &\leq -\alpha_0\left[1-\widetilde{F}_{0k}(\tau)-\widetilde{F}_{1k}(\tau)\right] \\
\left[1-\widetilde{F}_{1k}(\tau)-(1-p_k)\right] &\geq \alpha_0\left[1-\widetilde{F}_{0k}(\tau)-\widetilde{F}_{1k}(\tau)\right]
\end{aligned}
$$

Therefore

$$
\alpha_0 \leq \frac{1-\widetilde{F}_{1k}(\tau)-(1-p_k)}{1-\widetilde{F}_{0k}(\tau)-\widetilde{F}_{1k}(\tau)} = p_k\left[\frac{1-F_{1k}(\tau)}{1-F_k(\tau)}\right] \tag{12}
$$

Putting Everything Together For all k we have

$$
\alpha_0 \leq p_k \min_{\tau}\left\{\left[\frac{F_{1k}(\tau)}{F_k(\tau)}\right] \wedge \left[\frac{1-F_{1k}(\tau)}{1-F_k(\tau)}\right]\right\} \leq p_k \tag{13}
$$

$$
\alpha_1 \leq (1-p_k) \min_{\tau}\left\{\left[\frac{F_{0k}(\tau)}{F_k(\tau)}\right] \wedge \left[\frac{1-F_{0k}(\tau)}{1-F_k(\tau)}\right]\right\} \leq (1-p_k) \tag{14}
$$

Note that these bounds can only improve upon those derived in the previous section since the ratio of CDFs tends to one as $\tau \to \infty$. To derive these tighter bounds we have made no assumption regarding the relationship between Z and the error term ε. These bounds use only the assumption that $\alpha_0 + \alpha_1 < 1$, and the assumption that T is conditionally independent of Z, Y given T^*. Notice that that the bounds are related. In particular,

$$
p_k\left[\frac{F_{1k}(\tau)}{F_k(\tau)}\right] = 1-(1-p_k)\left[\frac{F_{0k}(\tau)}{F_k(\tau)}\right]
$$

and

$$
p_k\left[\frac{1-F_{1k}(\tau)}{1-F_k(\tau)}\right] = 1-(1-p_k)\left[\frac{1-F_{0k}(\tau)}{1-F_k(\tau)}\right]
$$

4 Even Stronger Bounds on α_0, α_1

Try applying the stochastic dominance conditions from our simulation study.

5 Independent Instrument

Assume that $Z \perp U$. The model is $Y = \beta T^* + U$ and

$$F_U(\tau) = P(U \leq \tau) = P(Y - \beta T^* \leq \tau)$$

but if Z is independent of U then it follows that

$$\begin{aligned} F_U(\tau) &= F_{U|Z=k}(\tau) = P(U \leq \tau | Z = k) = P(Y - \beta T^* \leq \tau | Z = k) \\ &= P(Y \leq \tau | T^* = 0, Z = k)(1 - p_k^*) + P(Y \leq \tau + \beta | T^* = 1, Z = k) p_k^* \\ &= (1 - p_k^*) F_{0k}^*(\tau) + p_k^* F_{1k}^*(\tau + \beta) \end{aligned}$$

for all k by the Law of Total Probability. Similarly,

$$F_k(\tau) = (1 - p_k^*) F_{0k}^*(\tau) + p_k^* F_{1k}^*(\tau)$$

and rearranging

$$(1 - p_k^*) F_{0k}^*(\tau) = F_k(\tau) - p_k^* F_{1k}^*(\tau)$$

Substituting this expression into the equation for $F_U(\tau)$ from above, we have

$$F_U(\tau) = F_k(\tau) + p_k^* \left[F_{1k}^*(\tau + \beta) - F_{1k}^*(\tau) \right]$$

for all k and all τ. Evaluating at two values k and ℓ in the support of Z and equating

$$F_k(\tau) + p_k^* \left[F_{1k}^*(\tau + \beta) - F_{1k}^*(\tau) \right] = F_\ell(\tau) + p_\ell^* \left[F_{1\ell}^*(\tau + \beta) - F_{1\ell}^*(\tau) \right]$$

or equivalently

$$F_k(\tau) - F_\ell(\tau) = p_\ell^* \left[F_{1\ell}^*(\tau + \beta) - F_{1\ell}^*(\tau) \right] - p_k^* \left[F_{1k}^*(\tau + \beta) - F_{1k}^*(\tau) \right] \tag{15}$$

for all τ. Now we simply need to re-express all of the "star" quantities, namely p_k^*, p_ℓ^* and $F_{1k}^*, F_{1\ell}^*$ in terms of α_0, α_1 and the *observable* probability distributions F_{1k} and $F_{1\ell}$ and observable probabilities p_k, p_ℓ. To do this, we use the fact that

$$\begin{aligned} F_{0k}(\tau) &= \frac{1 - \alpha_0}{1 - p_k}(1 - p_k^*) F_{0k}^*(\tau) + \frac{\alpha_1}{1 - p_k} p_k^* F_{1k}^*(\tau) \\ F_{1k}(\tau) &= \frac{\alpha_0}{p_k}(1 - p_k^*) F_{0k}^*(\tau) + \frac{1 - \alpha_1}{p_k} p_k^* F_{1k}^*(\tau) \end{aligned}$$

for all k by Bayes' rule. Solving these equations,

$$p_k^* F_{1k}^*(\tau) = \frac{1-\alpha_0}{1-\alpha_0-\alpha_1} p_k F_{1k}(\tau) - \frac{\alpha_0}{1-\alpha_0-\alpha_1}(1-p_k)F_{0k}(\tau)$$

for all k. Combining this with Equation 15, we find that

$$\begin{aligned}(1-\alpha_0-\alpha_1)\left[F_k(\tau)-F_\ell(\tau)\right] = \alpha_0 \left\{(1-p_k)\left[F_{0k}(\tau+\beta)-F_{0k}(\tau)\right]-(1-p_\ell)\left[F_{0\ell}(\tau+\beta)-F_{0\ell}(\tau)\right]\right\} \\ -(1-\alpha_0)\left\{p_k\left[F_{1k}(\tau+\beta)-F_{1k}(\tau)\right]-p_\ell\left[F_{1\ell}(\tau+\beta)-F_{1\ell}(\tau)\right]\right\}\end{aligned}$$

Now, define

$$\Delta_{tk}^{\tau}(\beta) = F_{tk}(\tau+\beta)-F_{tk}(\tau) = E\left[\frac{\mathbf{1}\left\{T=t,Z=k\right\}}{p_{tk}}\left(\mathbf{1}\left\{Y\leq\tau+\beta\right\}-\mathbf{1}\left\{Y\leq\tau\right\}\right)\right]$$

and note that we can express $F_k(\tau)-F_\ell(\tau)$ similarly as

$$F_k(\tau)-F_\ell(\tau) = E\left[\mathbf{1}\left\{Y\leq\tau\right\}\left(\frac{\mathbf{1}\left\{Z=k\right\}}{q_k}-\frac{\mathbf{1}\left\{Z=\ell\right\}}{q_\ell}\right)\right]$$

Using this notation, we can write the preceding as

$$(1-\alpha_0-\alpha_1)\left[F_k(\tau)-F_\ell(\tau)\right] = \alpha_0\left[(1-p_k)\Delta_{0k}^{\tau}(\beta)-(1-p_\ell)\Delta_{0\ell}^{\tau}(\beta)\right]-(1-\alpha_0)\left[p_k\Delta_{1k}^{\tau}(\beta)-p_\ell\Delta_{1\ell}^{\tau}(\beta)\right]$$

or in moment-condition form

$$\begin{aligned}E\Bigg[&(1-\alpha_0-\alpha_1)\mathbf{1}\left\{Y\leq\tau\right\}\left(\frac{\mathbf{1}\left\{Z=k\right\}}{q_k}-\frac{\mathbf{1}\left\{Z=\ell\right\}}{q_\ell}\right)-\left(\mathbf{1}\left\{Y\leq\tau+\beta\right\}-\mathbf{1}\left\{Y\leq\tau\right\}\right)\Bigg\{ \\ &\alpha_0\left((1-p_k)\frac{\mathbf{1}\left\{T=0,Z=k\right\}}{p_{0k}}-(1-p_\ell)\frac{\mathbf{1}\left\{T=0,Z=\ell\right\}}{p_{0\ell}}\right) \\ &-(1-\alpha_0)\left(p_k\frac{\mathbf{1}\left\{T=1,Z=k\right\}}{p_{1k}}-p_\ell\frac{\mathbf{1}\left\{T=1,Z=\ell\right\}}{p_{1\ell}}\right)\Bigg\}\Bigg]=0\end{aligned}$$

Each value of τ yields a moment condition.

6 Special Case: $\alpha_0 = 0$

In this case the expressions from above simplify to

$$(1-\alpha_1)\left[F_k(\tau)-F_\ell(\tau)\right] = \left[p_\ell F_{1\ell}(\tau+\beta)-p_kF_{1k}(\tau+\beta)-p_\ell F_{1\ell}(\tau)+p_kF_{1k}(\tau)\right] \tag{16}$$

for all τ. Now, provided that all of the CDFs are differentiable we have[1]

$$e^{i\omega\tau}(1-\alpha_1)\left[f_k(\tau)-f_\ell(\tau)\right] = e^{i\omega\tau}\left[p_\ell f_{1\ell}(\tau+\beta) - p_k f_{1k}(\tau+\beta) - p_\ell f_{1\ell}(\tau) + p_k f_{1k}(\tau)\right]$$

where we have pre-multiplied both sides by $e^{i\omega\tau}$. Finally, integrating both sides with respect to τ over $(-\infty,\infty)$, we have

$$(1-\alpha_1)\left[\varphi_k(\omega)-\varphi_\ell(\omega)\right] = \left\{\int_{-\infty}^{\infty} e^{i\omega\tau}\left[p_\ell f_{1\ell}(\tau+\beta) - p_k f_{1k}(\tau+\beta)\right]\,d\tau - p_\ell\varphi_{1\ell}(\omega) + p_k\varphi_{1k}(\omega)\right\}$$

where φ_k is the conditional characteristic function of Y given $Z=k$ and φ_{1k} is the conditional characteristic function of Y given $T=1, Z=k$. Finally,

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{i\omega\tau} p_\ell f_{1\ell}(\tau+\beta)\,d\tau &= e^{i\omega\beta} p_\ell \int_{u=-\infty+\beta}^{u=\infty+\beta} e^{i\omega u} f_{1\ell}(u)\,du \\
&= e^{-i\omega\beta} p_\ell \varphi_{1\ell}(\omega)
\end{aligned}$$

using the substitution $u=\tau+\beta$. Changing subscripts, the same holds for k and thus

$$(1-\alpha_1)\left[\varphi_k(\omega)-\varphi_\ell(\omega)\right] = e^{-i\omega\beta}\left[p_\ell\varphi_{1\ell}(\omega) - p_k\varphi_{1k}(\omega)\right] + \left[p_k\varphi_{1k}(\omega) - p_\ell\varphi_{1\ell}(\omega)\right]$$

which, after collecting terms, simplifies to

$$(1-\alpha_1)\left[\varphi_k(\omega)-\varphi_\ell(\omega)\right] = \left(e^{-i\omega\beta}-1\right)\left[p_\ell\varphi_{1\ell}(\omega) - p_k\varphi_{1k}(\omega)\right] \tag{17}$$

for all ω. Equation 17 contains exactly the same information as Equation 16 but gives us a more convenient way to prove identification since β enters in a simpler way. Leibniz's formula for the rth derivative of a product of two functions f and g is:

$$(fg)^{(r)} = \sum_{s=0}^{r}\binom{r}{s} f^{(s)} g^{(r-s)}$$

[1]There must be a way to generalize this using Lebesgue.

where $f^{(r)}$ denotes the rth derivative of the function f and $g^{(r-s)}$ denotes the $(r-s)$th derivative of the function g. Applying this to the RHS, $R(\omega)$ of Equation 17 gives

$$\begin{aligned}\frac{d}{d\omega^r}R(\omega) &= \sum_{s=0}^{r}\binom{r}{s}\frac{d}{d\omega^s}\left(e^{-i\omega\beta}-1\right)\frac{d}{d\omega^{r-s}}\left[p_\ell\varphi_{1\ell}(\omega)-p_k\varphi_{1k}(\omega)\right]\\ &= \left(e^{-i\omega\beta}-1\right)\left[p_\ell\varphi_{1\ell}^{(r)}(\omega)-p_k\varphi_{1k}^{r}(\varphi)\right]+e^{-i\omega\beta}\sum_{s=1}^{r}\binom{r}{s}(-i\beta)^s\left[p_\ell\varphi_{1\ell}^{(r-s)}(\omega)-p_k\varphi_{1k}^{(r-s)}(\omega)\right]\end{aligned}$$

where we split off the $s=0$ term because our generic expression for the sth derivative of $(e^{-i\omega\beta}-1)$ only applies for $s\geq 1$. Evaluating at zero:

$$\frac{d}{d\omega^r}R(0)=\sum_{s=1}^{r}\binom{r}{s}(-i\beta)^s\left[p_\ell\varphi_{1\ell}^{(r-s)}(0)-p_k\varphi_{1k}^{(r-s)}(0)\right]$$

Combining this with the LHS of Equation 17, also differentiated r times and evaluated at zero, we have

$$(1-\alpha_1)\left[\varphi_k^{(r)}(0)-\varphi_\ell^{(r)}(0)\right]=\sum_{s=1}^{r}\binom{r}{s}(-i\beta)^s\left[p_\ell\varphi_{1\ell}^{(r-s)}(0)-p_k\varphi_{1k}^{(r-s)}(0)\right]$$

Now, recall that if $\varphi(\omega)$ is the characteristic function of Y then $\varphi^{(r)}(0)=i^rE[Y^r]$ provided that the expectation exists where $\varphi^{(r)}$ denotes the rth derivative of φ. The same applies for the conditional characteristic functions we consider here. Hence, provided that the rth moments exist,

$$i^r(1-\alpha_1)\left\{E[Y^r|Z=k]-E[Y^r|Z=\ell]\right\}=\sum_{s=1}^{r}\binom{r}{s}(-i\beta)^s i^{r-s}\left(p_\ell E\left[Y^{r-s}|T=1,Z=\ell\right]-p_kE\left[Y^{r-s}|T=1,Z=k\right]\right)$$

After simplifying the terms involving i and cancelling them from both sides,

$$(1-\alpha_1)\left(E[Y^r|Z=k]-E[Y^r|Z=\ell]\right)=\sum_{s=1}^{r}\binom{r}{s}(-\beta)^s\left(p_\ell E\left[Y^{r-s}|T=1,Z=\ell\right]-p_kE\left[Y^{r-s}|T=1,Z=k\right]\right)$$

again provided that the moments exist. Abbreviating the conditional expectations according to $E[Y^r|Z=k]=E_k[Y^r]$ and $E[Y^r|T=t,Z=k]=E_{tk}[Y^r]$, this becomes

$$(1-\alpha_1)\left(E_k[Y^r]-E_\ell[Y^r]\right)=\sum_{s=1}^{r}\binom{r}{s}(-\beta)^s\left(p_\ell E_{1\ell}\left[Y^{r-s}\right]-p_kE_{1k}\left[Y^{r-s}\right]\right) \qquad (18)$$

Equation 18 can be used to generate moment equations that are implied by the Equation 17 and the equivalent representation in terms of CDFs: Equation 16. Assuming that the conditional first moments exist, we can evaluate Equation 18 at $r = 1$, yielding

$$\begin{aligned}(1-\alpha_1)\left(E_k[Y]-E_\ell[Y]\right) &= \sum_{s=1}^{1}\binom{1}{s}(-\beta)^s\left(p_\ell E_{1\ell}\left[Y^{1-s}\right]-p_k E_{1k}\left[Y^{1-s}\right]\right)\\ &= -\beta\left(p_\ell - p_k\right)\end{aligned}$$

Rearranging, this gives us the expression for the probability limit of the Wald estimator

$$\mathcal{W} \equiv \frac{E_k[Y]-E_\ell[Y]}{p_k - p_\ell} = \frac{\beta}{1-\alpha_1} \tag{19}$$

Evaluating Equation 18 at $r = 2$, we have

$$\begin{aligned}(1-\alpha_1)\left(E_k[Y^2]-E_\ell[Y^2]\right) &= \sum_{s=1}^{2}\binom{2}{s}(-\beta)^s\left(p_\ell E_{1\ell}\left[Y^{2-s}\right]-p_k E_{1k}\left[Y^{2-s}\right]\right)\\ &= 2\beta\left(p_k E_{1k}[Y]-p_\ell E_{1\ell}[Y]\right)-\beta^2\left(p_k-p_\ell\right)\end{aligned}$$

Rearranging, we have

$$E_k[Y^2]-E_\ell[Y^2] = \frac{\beta}{1-\alpha_1}\left[2\left(p_k E_{1k}[Y]-p_\ell E_{1\ell}[Y]\right)-\beta(p_k-p_\ell)\right] \tag{20}$$

Substituting Equation 19, we can replace $\beta/(1-\alpha_1)$ with a function of observables only, namely $\mathcal{W}$. Solving, we find that

$$\beta = \frac{2\left(p_k E_{1k}[Y]-p_\ell E_{1\ell}[Y]\right)}{p_k-p_\ell} - \frac{E_k[Y^2]-E_\ell[Y^2]}{E_k[Y]-E_\ell[Y]} \tag{21}$$

This allows us to state low-level sufficient conditions for identification:

(a) $\alpha_1 < 1$

(b) $p_k \neq p_\ell$

(c) $E_k[Y] \neq E_\ell[Y]$

(d) $E_{1k}[|Y|], E_{1\ell}[|Y|], E_k[|Y^2|], E_\ell[|Y^2|] < \infty$.

Note that, although $\beta = 0$ is always a solution of Equation 16 this solution is ruled out by the assumption that $E_k[Y] \neq E_\ell[Y]$ via Equation 19. The mis-classification error rate α_1 is likewise uniquely identified under these assumptions. Substituting $\beta/\mathcal{W} = 1 - \alpha_1$ into

Equation 21

$$
\begin{aligned}
(1-\alpha_1) &= \left\{\frac{p_k - p_\ell}{E_k[Y]-E_\ell[Y]}\right\}\left\{\frac{2\left(p_k E_{1k}[Y] - p_\ell E_{1\ell}[Y]\right)}{p_k - p_\ell} - \frac{E_k[Y^2]-E_\ell[Y^2]}{E_k[Y]-E_\ell[Y]}\right\}\\
&= \frac{2\left(p_k E_{1k}[Y] - p_\ell E_{1\ell}[Y]\right)}{E_k[Y]-E_\ell[Y]} - (p_k - p_\ell)\left\{\frac{E_k[Y^2]-E_\ell[Y^2]}{\left(E_k[Y]-E_\ell[Y]\right)^2}\right\}
\end{aligned}
$$

and thus

$$
\alpha_1 = 1 + (p_k - p_\ell)\left\{\frac{E_k[Y^2]-E_\ell[Y^2]}{\left(E_k[Y]-E_\ell[Y]\right)^2}\right\} - \frac{2\left(p_k E_{1k}[Y] - p_\ell E_{1\ell}[Y]\right)}{E_k[Y]-E_\ell[Y]}
$$

7 Identification in the General Case

8 Characteristic Functions

Recall from above that in the general case an independent instrument combined with non-differential measurement error implies that

$$
\begin{aligned}
(1-\alpha_0-\alpha_1)\left[F_k(\tau)-F_\ell(\tau)\right] &= \alpha_0\left\{(1-p_k)\left[F_{0k}(\tau+\beta)-F_{0k}(\tau)\right]-(1-p_\ell)\left[F_{0\ell}(\tau+\beta)-F_{0\ell}(\tau)\right]\right\}\\
&\quad -(1-\alpha_0)\left\{p_k\left[F_{1k}(\tau+\beta)-F_{1k}(\tau)\right]-p_\ell\left[F_{1\ell}(\tau+\beta)-F_{1\ell}(\tau)\right]\right\}
\end{aligned}
$$

Using the same steps as in the preceding section, we can convert this expression into characteristic function form by differentiating each side, multiplying by $e^{i\omega\tau}$ and then integrating with respect to τ, yielding

$$
\begin{aligned}
(1-\alpha_0-\alpha_1)\left[\varphi_k(\omega)-\varphi_\ell(\omega)\right] &= \alpha_0\left\{(1-p_k)\left(e^{-i\omega\beta}-1\right)\varphi_{0k}(\omega)-(1-p_\ell)\left(e^{-i\omega\beta}-1\right)\varphi_{0\ell}(\omega)\right\}\\
&\quad -(1-\alpha_0)\left\{p_k\left(e^{-i\omega\beta}-1\right)\varphi_{1k}(\omega)-p_\ell\left(e^{-i\omega\beta}-1\right)\varphi_{1\ell}(\omega)\right\}
\end{aligned}
$$

which simplifies to

$$
\varphi_k(\omega)-\varphi_\ell(\omega) = \left(e^{-i\omega\beta}-1\right)\left(\frac{\alpha_0\left[(1-p_k)\varphi_{0k}(\omega)-(1-p_\ell)\varphi_{0\ell}(\omega)\right]-(1-\alpha_0)\left[p_k\varphi_{1k}(\omega)-p_\ell\varphi_{1\ell}(\omega)\right]}{1-\alpha_0-\alpha_1}\right).
$$

As above, we will differentiate both sides of this expression r times and evaluate at $\omega = 0$. Steps nearly identical to those given above yield

$$
\begin{aligned}
(1-\alpha_0-\alpha_1)\left(E_k[Y^r]-E_\ell[Y^r]\right) = \alpha_0 \sum_{s=1}^{r}\binom{r}{s}(-\beta)^s\left\{(1-p_k)E_{0k}[Y^{r-s}]-(1-p_\ell)E_{0\ell}[Y^{r-s}]\right\} \\
-(1-\alpha_0)\sum_{s=1}^{r}\binom{r}{s}(-\beta)^s\left\{p_kE_{1k}[Y^{r-s}]-p_\ell E_{1\ell}[Y^{r-s}]\right\}
\end{aligned}
$$

First Moments Taking $r = 1$ gives

$$
(1-\alpha_0-\alpha_1)\left(E_k[Y]-E_\ell[Y]\right) = \beta(p_k-p_\ell)
$$

Simplifying,

$$
\mathcal{W} \equiv \frac{E_k[Y]-E_\ell[Y]}{p_k-p_\ell} = \frac{\beta}{1-\alpha_0-\alpha_1} \tag{22}
$$

Second Moments Now, taking $r = 2$ gives

$$
\begin{aligned}
(1-\alpha_0-\alpha_1)\left(E_k[Y^2]-E_\ell[Y^2]\right) &= \alpha_0\left\{[(1-p_k)E_{0k}[Y]-(1-p_\ell)E_{0\ell}]-\beta^2(p_k-p_\ell)\right\} \\
&\quad -(1-\alpha_0)\left\{-2\beta\left(p_kE_{1k}[Y]-p_\ell E_{1\ell}[Y]\right)+\beta^2(p_k-p_\ell)\right\} \\
&= -2\beta\alpha_0\left\{(1-p_k)E_{0k}[Y]-(1-p_\ell)E_{0\ell}[Y]p_kE_{1k}[Y]+p_\ell E_{1\ell}[Y]\right\} \\
&\quad +2\beta\left(p_kE_{1k}[Y]-p_\ell E_{1\ell}[Y]\right)-(p_k-p_\ell)\beta^2(\alpha_0+1-\alpha_0) \\
&= -2\beta\left\{\alpha_0\left(E_k[Y]-E_\ell[Y]\right)-\left(p_kE_{1k}[Y]-p_\ell E_{1\ell}[Y]\right)\right\}-\beta^2(p_k-p_\ell)
\end{aligned}
$$

Now, simplifying

$$
(1-\alpha_0-\alpha_1)\left(\frac{E_k[Y^2]-E_\ell[Y^2]}{p_k-p_\ell}\right) = -2\beta\alpha_0\left(\frac{E_k[Y]-E_k[Y]}{p_k-p_\ell}\right)+2\beta\left(\frac{p_{1k}E_{1k}[Y]-p_\ell E_{1\ell}[Y]}{p_k-p_\ell}\right)-\beta^2
$$

and substituting Equation 22 to eliminate β, this becomes

$$
\begin{aligned}
(1-\alpha_0-\alpha_1)\left(\frac{E_k[Y^2]-E_\ell[Y^2]}{p_k-p_\ell}\right) &= -2\alpha_0(1-\alpha_0-\alpha_1)\mathcal{W}^2+2\mathcal{W}(1-\alpha_0-\alpha_1)\left(\frac{p_{1k}E_{1k}[Y]-p_\ell E_{1\ell}[Y]}{p_k-p_\ell}\right) \\
&\quad -(1-\alpha_0-\alpha_1)^2\mathcal{W}^2 \\
\left(\frac{E_k[Y^2]-E_\ell[Y^2]}{p_k-p_\ell}\right) &= -2\alpha_0\mathcal{W}^2+2\mathcal{W}\left(\frac{p_{1k}E_{1k}[Y]-p_\ell E_{1\ell}[Y]}{p_k-p_\ell}\right)-(1-\alpha_0-\alpha_1)\mathcal{W}^2
\end{aligned}
$$

And thus, simplifying

$$
\begin{aligned}
-2\alpha_0\mathcal{W}^2 - (1-\alpha_0-\alpha_1)\mathcal{W}^2 &= \left(\frac{E_k[Y^2]-E_\ell[Y^2]}{p_k-p_\ell}\right) - 2\mathcal{W}\left(\frac{p_{1k}E_{1k}[Y]-p_\ell E_{1\ell}[Y]}{p_k-p_\ell}\right)\\
\alpha_1-\alpha_0 &= 1+\left[\frac{E_k[Y^2]-E_\ell[Y^2]}{\mathcal{W}^2(p_k-p_\ell)}\right] - 2\left[\frac{p_{1k}E_{1k}[Y]-p_\ell E_{1\ell}[Y]}{\mathcal{W}(p_k-p_\ell)}\right]
\end{aligned}
$$

and therefore

$$
\alpha_1-\alpha_0 = 1+(p_k-p_\ell)\left[\frac{E_k[Y^2]-E_\ell[Y^2]}{(E_k[Y]-E_\ell[Y])^2}\right] - 2\left[\frac{p_{1k}E_{1k}[Y]-p_\ell E_{1\ell}[Y]}{E_k[Y]-E_\ell[Y]}\right] \tag{23}
$$

"Product" Moments Recall that in our initial draft of the paper we worked with moments such as $E[TY|Z=k], E[TY|Z=\ell]$ and $E[TY^2|Z=k], E[TY^2|Z=\ell]$. In the notation of this document, we can express these quantities as follows:

$$
\begin{aligned}
E[TY^r|z=k] &= E[TY^r|T=1,z=k]p_k + E[TY^r|T=0,z=k](1-p_k)\\
&= p_kE[Y^r|T=1,z=k]+0\\
&= p_kE_{1k}[Y^r]
\end{aligned}
$$

for any r. We will use this relationship to motivate some shorthand notation below.

Some Shorthand The notation above is becoming very cumbersome and we haven't even looked at the third moments yet! To make life easier, define the following:

$$
\begin{aligned}
\widetilde{y^r_{1k}} &= p_kE_{1k}[Y^r]\\
\widetilde{y^r_{0k}} &= (1-p_k)E_{1k}[Y^r]\\
\Delta\overline{y^r} &= E_k[Y^r]-E_\ell[Y^r]\\
\Delta\overline{Ty^r} &= p_kE_{1k}[Y^r]-p_\ell E_{1\ell}[Y^r] = \widetilde{y^r_{1k}} - \widetilde{y^r_{1k}}\\
\mathcal{W} &= (E_k[Y]-E_\ell[Y])/(p_k-p_\ell)
\end{aligned}
$$

for all r. When no r superscript is given this means $r=1$. Note, moreover, that when $r=0$ we have $\widetilde{y^0_{1k}} = p_k$ and $\widetilde{y^0_{0k}} = (1-p_k)$. Thus $\Delta\overline{Ty^0} = p_k - p_\ell$. In contrast, $\Delta y^0 = 0$.

Among other things, this notation will make it easier for us to link the derivations here to our earlier derivations from the first draft of the paper that used slightly different notation and did not work explicitly with the independence of the instrument.

Simplifying the Moment Equalities Using the final two pieces of notation defined in the preceding section, we can re-rewrite the collection of moment equalities arising from the characteristic function equations as

$$(1-\alpha_0-\alpha_1)\Delta\overline{y^r} = \sum_{s=1}^{r}\binom{r}{s}(-\beta)^s\left[\alpha_0\left(\widetilde{y_{0k}^{r-s}}-\widetilde{y_{0\ell}^{r-s}}\right)-(1-\alpha_0)\left(\widetilde{y_{1k}^{r-s}}-\widetilde{y_{1\ell}^{r-s}}\right)\right]$$

Now, simplifying the terms in the square brackets,

$$\begin{aligned}\alpha_0\left(\widetilde{y_{0k}^{r-s}}-\widetilde{y_{0\ell}^{r-s}}\right)-(1-\alpha_0)\left(\widetilde{y_{1k}^{r-s}}-\widetilde{y_{1\ell}^{r-s}}\right) &= \alpha_0\left[\left(\widetilde{y_{0k}^{r-s}}+\widetilde{y_{1k}^{r-s}}\right)-\left(\widetilde{y_{0\ell}^{r-s}}+\widetilde{y_{1\ell}^{r-s}}\right)\right]-\left(\widetilde{y_{1k}^{r-s}}-\widetilde{y_{1\ell}^{r-s}}\right)\\ &= \alpha_0\left(E_k[Y^{r-s}]-E_\ell[Y^{r-s}]\right)-\Delta\overline{Ty^{r-s}}\\ &= \alpha_0\Delta\overline{y^{r-s}}-\Delta\overline{Ty^{r-s}}\end{aligned}$$

and hence

$$(1-\alpha_0-\alpha_1)\Delta\overline{y^r} = \sum_{s=1}^{r}\binom{r}{s}(-\beta)^s\left(\alpha_0\Delta\overline{y^{r-s}}-\Delta\overline{Ty^{r-s}}\right) \tag{24}$$

Third Moments Evaluating Equation 24 at $r=3$

$$\begin{aligned}(1-\alpha_0-\alpha_1)\Delta\overline{y^3} &= \sum_{s=1}^{3}\binom{3}{s}(-\beta)^s\left(\alpha_0\Delta\overline{y^{3-s}}-\Delta\overline{Ty^{3-s}}\right)\\ &= -3\beta\left(\alpha_0\Delta\overline{y^2}-\Delta\overline{Ty^2}\right)+3\beta^2\left(\alpha_0\Delta\overline{y}-\Delta\overline{Ty}\right)+\beta^3(p_k-p_\ell)\end{aligned}$$

Solving the System Using $\mathcal{W}=\beta/(1-\alpha_0-\alpha_1)$ we can re-write the third moment expression as follows

$$\begin{aligned}\Delta\overline{y^3} &= -3\mathcal{W}\left(\alpha_0\Delta\overline{y^2}-\Delta\overline{Ty^2}\right)+3\beta\mathcal{W}\left(\alpha_0\Delta\overline{y}-\Delta\overline{Ty}\right)+\beta^2\mathcal{W}(p_k-p_\ell)\\ \frac{\Delta\overline{y^3}}{\mathcal{W}(p_k-p_\ell)} &= \beta^2+3\beta\left(\frac{\alpha_0\Delta\overline{y}-\Delta\overline{Ty}}{p_k-p_\ell}\right)-3\left(\frac{\alpha_0\Delta\overline{y^2}-\Delta\overline{Ty^2}}{p_k-p_\ell}\right)\\ \frac{\Delta\overline{y^3}-3\mathcal{W}\Delta\overline{y^2T}}{\mathcal{W}(p_k-p_\ell)} &= \beta^2+3\beta\left(\frac{\alpha_0\Delta\overline{y}-\Delta\overline{Ty}}{p_k-p_\ell}\right)-3\left(\frac{\alpha_0\Delta\overline{y^2}}{p_k-p_\ell}\right)\end{aligned}$$

Now, translating the second moment equation into the shorthand notation defined above, we have

Simplifying the Characteristic Function Equation From above, we have

$$\varphi_k(\omega) - \varphi_\ell(\omega) = \left(e^{-i\omega\beta} - 1\right)\left(\frac{\alpha_0\left[(1-p_k)\varphi_{0k}(\omega) - (1-p_\ell)\varphi_{0\ell}(\omega)\right] - (1-\alpha_0)\left[p_k\varphi_{1k}(\omega) - p_\ell\varphi_{1\ell}(\omega)\right]}{1-\alpha_0-\alpha_1}\right).$$

Using the fact that $\varphi_k = p_k\varphi_{1k} + (1-p_k)\varphi_{0k}$, we can simplify this further, yielding

$$(1-\alpha_0-\alpha_1) = \left(e^{-i\omega\beta} - 1\right)\left[\alpha_0 - \xi(\omega)\right]$$

where we define

$$\xi(\omega) \equiv \frac{\varphi_k(\omega) - \varphi_\ell(\omega)}{p_k\varphi_{1k}(\omega) - p_\ell\varphi_{1\ell}(\omega)}$$

Now, re-arranging

$$(1-\alpha_1) - \xi(\omega) = e^{-i\omega\beta}\left[\alpha_0 - \xi(\omega)\right]$$

or equivalently

$$e^{i\omega\beta}\left[(1-\alpha_1) - \xi(\omega)\right] = \alpha_0 - \xi(\omega)$$

or

$$e^{i\omega\beta} = \frac{\alpha_0 - \xi(\omega)}{(1-\alpha_1) - \xi(\omega)}$$

provided the denominator does not vanish. By taking differences or ratios evaluated at ω_1 and ω_2 we can eliminate β, α_0 or α_1 but it's not clear how or if we can prove identification in terms of a restriction on the characteristic functions.

Suppose we consider three values ω_1, ω_2 and ω_3 for which that yield to distinct, non-zero values ξ_1, ξ_2 and ξ_3 of $\xi(\omega)$.

$$e^{i\omega_1\beta}\left[(1-\alpha_1) - \xi_1\right] - e^{i\omega_2\beta}\left[(1-\alpha_1) - \xi_2\right] = \xi_2 - \xi_1$$

8.1 Simplifying the Characteristic CDF Equation

Recall from above that

$$\begin{aligned}(1-\alpha_0-\alpha_1)\left[F_k(\tau) - F_\ell(\tau)\right] &= \alpha_0\left\{(1-p_k)\left[F_{0k}(\tau+\beta) - F_{0k}(\tau)\right] - (1-p_\ell)\left[F_{0\ell}(\tau+\beta) - F_{0\ell}(\tau)\right]\right\} \\ &\quad - (1-\alpha_0)\left\{p_k\left[F_{1k}(\tau+\beta) - F_{1k}(\tau)\right] - p_\ell\left[F_{1\ell}(\tau+\beta) - F_{1\ell}(\tau)\right]\right\}\end{aligned}$$

We can simplify the RHS as follows

$$\begin{aligned}\text{RHS} &= \alpha_0\left\{\left[F_k(\tau+\beta) - F_\ell(\tau+\beta)\right] - \left[F_k(\tau) - F_\ell(\tau)\right]\right\} \\ &\quad - \left\{\left[p_kF_{1k}(\tau+\beta) - p_\ell F_{1\ell}(\tau+\beta)\right] - \left[p_kF_{1k}(\tau) - p_\ell F_{1\ell}(\tau)\right]\right\}\end{aligned}$$

Now, define

$$\Delta(\tau) = F_k(\tau) - F_\ell(\tau)$$
$$\widetilde{\Delta}_1(\tau) = p_k F_{1k}(\tau) - p_\ell F_{1\ell}(\tau)$$

Using this notation, our equation becomes

$$(1 - \alpha_0 - \alpha_1)\Delta(\tau) = \alpha_0 \left[\Delta(\tau+\beta) - \Delta(\tau)\right] - \left[\widetilde{\Delta}_1(\tau+\beta) - \widetilde{\Delta}_1(\tau)\right]$$

which simplifies to

$$\widetilde{\Delta}_1(\tau+\beta) - \widetilde{\Delta}_1(\tau) = \alpha_0\Delta(\tau+\beta) - (1-\alpha_1)\Delta(\tau)$$

Suppose $\alpha_0 = 0$: In this case we obtain

$$(1-\alpha_1) = \frac{\widetilde{\Delta}_1(\tau) - \widetilde{\Delta}_1(\tau+\beta)}{\Delta(\tau)}$$

Now, evaluating at two values of τ and taking differences, we find

$$\frac{\widetilde{\Delta}_1(\tau) - \widetilde{\Delta}_1(\tau+\beta)}{\Delta(\tau)} - \frac{\widetilde{\Delta}_1(\tau') - \widetilde{\Delta}_1(\tau'+\beta)}{\Delta(\tau')} = 0$$

Suppose $\alpha_1 = 0$: In this case we obtain

$$\alpha_0 = \frac{\widetilde{\Delta}_1(\tau+\beta) - \widetilde{\Delta}_1(\tau) + \Delta(\tau)}{\Delta(\tau+\beta)}$$

Again, taking differences evaluated at two values of τ,

$$\frac{\widetilde{\Delta}_1(\tau+\beta) - \widetilde{\Delta}_1(\tau) + \Delta(\tau)}{\Delta(\tau+\beta)} - \frac{\widetilde{\Delta}_1(\tau'+\beta) - \widetilde{\Delta}_1(\tau') + \Delta(\tau')}{\Delta(\tau'+\beta)} = 0$$

Some Equations to Check Numerically We can use the same basic idea when either α_0 or α_1 is known but nonzero. This isn't realistic in practice, but can be used to check our equations:

$$\alpha_0 = \frac{\widetilde{\Delta}_1(\tau+\beta) - \widetilde{\Delta}_1(\tau) + (1-\alpha_1)\Delta(\tau)}{\Delta(\tau+\beta)}$$

$$(1-\alpha_1) = \frac{\alpha_0\Delta(\tau+\beta) + \widetilde{\Delta}_1(\tau) - \widetilde{\Delta}_1(\tau+\beta)}{\Delta(\tau)}$$

As above, after substituting the true value of either α_1 or α_0, we can eliminate the remaining mis-classification probability by evaluating at two quantiles τ, τ' and taking differences.

These appear to work just fine!

What if α_0 and α_1 are both unknown? Suppose we take differences at two quantiles τ and ν to eliminate α_1:

$$\left[\frac{\alpha_0\Delta(\tau+\beta) + \widetilde{\Delta}_1(\tau) - \widetilde{\Delta}_1(\tau+\beta)}{\Delta(\tau)}\right] - \left[\frac{\alpha_0\Delta(\nu+\beta) + \widetilde{\Delta}_1(\nu) - \widetilde{\Delta}_1(\nu+\beta)}{\Delta(\nu)}\right] = 0$$

$$\alpha_0\left[\frac{\Delta(\tau+\beta)}{\Delta(\tau)} - \frac{\Delta(\nu+\beta)}{\Delta(\nu)}\right] - \left[\frac{\widetilde{\Delta}_1(\tau) - \widetilde{\Delta}_1(\tau+\beta)}{\Delta(\tau)} - \frac{\widetilde{\Delta}_1(\nu) - \widetilde{\Delta}_1(\nu+\beta)}{\Delta(\nu)}\right] = 0$$

The Equation that Didn't Work...

$$\frac{[\widetilde{\Delta}_1(\tau+\beta) - \widetilde{\Delta}_1(\tau)] - [\widetilde{\Delta}_1(\tau'+\beta) - \widetilde{\Delta}_1(\tau')]}{\Delta(\tau+\beta) - \Delta(\tau'+\beta)} - \frac{[\widetilde{\Delta}_1(\nu+\beta) - \widetilde{\Delta}_1(\nu)] - [\widetilde{\Delta}_1(\nu'+\beta) - \widetilde{\Delta}_1(\nu')]}{\Delta(\nu+\beta) - \Delta(\nu'+\beta)} = 0$$

where $\Delta(\nu) = \Delta(\nu')$ and $\Delta(\tau) = \Delta(\tau')$.

9 New Results from September 2016

9.0.1 Relationship between observed and unobserved CDFs

Let

$$F^*_{tk}(\tau) = P(Y \leq \tau | T^* = t, z_k)$$
$$F_{tk}(\tau) = P(Y \leq \tau | T = t, z_k)$$

Now, by the assumption of non-differential measurement error,

$$\begin{aligned} p_k F_{1k}(\tau) &= (1-\alpha_1)p_k^* F_{1k}^*(\tau) + \alpha_0(1-p_k^*)F_{0k}^*(\tau) \\ (1-p_k)F_{0k}(\tau) &= \alpha_1 p_k^* F_{1k}^*(\tau) + (1-\alpha_0)(1-p_k^*)F_{0k}^*(\tau) \end{aligned}$$

Solving the linear system as above, we find that

$$\begin{aligned} F_{0k}^*(\tau) &= F_{0k}(\tau) + \left(\frac{\alpha_1 p_k}{1-p_k-\alpha_1}\right)[F_{0k}(\tau) - F_{1k}(\tau)] \\ F_{1k}^*(\tau) &= F_{1k}(\tau) + \left(\frac{\alpha_0(1-p_k)}{p_k-\alpha_0}\right)[F_{1k}(\tau) - F_{0k}(\tau)] \end{aligned}$$

9.1 Can we relax the measurement error assumptions?

Suppose that we continue to assume that $P(Y|T^*, T, z) = P(Y|T^*, z)$ but relax the assumption that $P(T|T^*, z) = P(T|T^*)$. Define:

$$\begin{aligned} \alpha_{0k} &= P\left(T=1|T^*=1, z_k\right) \\ \alpha_{1k} &= P\left(T=1|T^*=0, z_k\right) \end{aligned}$$

As before, the Wald estimator converges in probability to

$$\mathcal{W} = \frac{E[Y|z_k] - E[Y|z_\ell]}{p_k - p_\ell}$$

but the relationship between $p_1 - p_0$ and the unobserved $p_1^* - p_0^*$ changes. By the law of total probability

$$\begin{aligned} p_k = P(T=1|z_k) &= P(T=1|T^*=1, z_k)P(T^*=1|z_k) + P(T=1|T^*=0, z_k)P(T^*=0|z_k) \\ &= (1-\alpha_{1k})p_k^* + \alpha_{0k}(1-p_k^*) = p_k^*(1-\alpha_{0k}-\alpha_{1k}) + \alpha_{0k} \end{aligned}$$

and thus

$$p_k^* = \frac{p_k - \alpha_{0k}}{1-\alpha_{0k}-\alpha_{1k}}, \quad 1-p_k^* = \frac{1-p_k-\alpha_{1k}}{1-\alpha_{0k}-\alpha_{1k}}.$$

Thus, we have

$$\begin{aligned}
p_k^* - p_\ell^* &= \left(\frac{p_k - \alpha_{0k}}{1 - \alpha_{0k} - \alpha_{1k}}\right) - \left(\frac{p_0 - \alpha_{0\ell}}{1 - \alpha_{0\ell} - \alpha_{1\ell}}\right) \\
&= \frac{(p_k - \alpha_{0k})(1 - \alpha_{0\ell} - \alpha_{1\ell}) - (p_0 - \alpha_{0\ell})(1 - \alpha_{0k} - \alpha_{1k})}{(1 - \alpha_{0k} - \alpha_{1k})(1 - \alpha_{0\ell} - \alpha_{1\ell})}
\end{aligned}$$

9.2 Is there a LATE interpretation of our results?

Let $J \in \{a, c, d, n\}$ index an individual's *type*: always-taker, complier, defier, or never-taker. Let $\pi_a, \pi_c, \pi_d, \pi_n$ denote the population proportions of always-takers, compliers, defiers, and never-takers. The unconfounded type assumption is $P(J = j|z = 1) = P(J = j|z = 0)$. Combined with the law of total probability, this gives

$$\begin{aligned}
p_1^* &= P(T^* = 1|z = 1) = \pi_a + \pi_c \\
1 - p_1^* &= P(T^* = 0|z = 1) = \pi_d + \pi_n \\
p_0^* &= P(T^* = 1|z = 0) = \pi_d + \pi_a \\
1 - p_0^* &= P(T^* = 0|z = 0) = \pi_n + \pi_c
\end{aligned}$$

Imposing no-defiers, $\pi_d = 0$, these expressions simplify to

$$\begin{aligned}
p_1^* &= \pi_a + \pi_c \\
1 - p_1^* &= \pi_n \\
p_0^* &= \pi_a \\
1 - p_0^* &= \pi_n + \pi_c
\end{aligned}$$

Solving for π_c, we see that

$$\begin{aligned}
\pi_c &= p_1^* - p_0^* \\
\pi_a &= p_0^* \\
\pi_n &= 1 - p_1^*
\end{aligned}$$

Now, let $Y(1)$ indicate the potential outcome when $T^* = 1$ and $Y(0)$ indicate the potential outcome when $T^* = 0$. The standard LATE assumptions (no defiers, mean exclusion, unconfounded type) imply

$$
\begin{aligned}
\mathbb{E}\left(Y|T^*=1,z=1\right) &= \left(\frac{p_0^*}{p_1^*}\right)\mathbb{E}\left[Y(1)|J=a\right]+\left(\frac{p_1^*-p_0^*}{p_1^*}\right)\mathbb{E}\left[Y(1)|J=c\right]\\
\mathbb{E}\left(Y|T^*=0,z=0\right) &= \left(\frac{p_1^*-p_0^*}{1-p_0^*}\right)\mathbb{E}\left[Y(0)|J=c\right]+\left(\frac{1-p_1^*}{1-p_0^*}\right)\mathbb{E}\left[Y(0)|J=n\right]\\
\mathbb{E}\left(Y|T^*=1,z=0\right) &= \mathbb{E}\left[Y(1)|J=a\right]\\
\mathbb{E}\left(Y|T^*=0,z=1\right) &= \mathbb{E}\left[Y(0)|J=n\right]
\end{aligned}
$$

9.2.1 LATE Version of Theorem 2 from the Draft

$$
\begin{aligned}
\Delta\overline{yT} &= \mathbb{E}\left(yT|z=1\right)-\mathbb{E}\left(yT|z=0\right)\\
&= (1-\alpha_1)\left[p_1^*\mathbb{E}\left(y|T^*=1,z=1\right)-p_0^*\mathbb{E}\left(y|T^*=1,z=0\right)\right]\\
&\qquad+\alpha_0\left[(1-p_1^*)\mathbb{E}\left(y|T^*=0,z=1\right)-(1-p_0^*)\mathbb{E}\left(y|T^*,z=0\right)\right]
\end{aligned}
$$

So we find that

$$
\begin{aligned}
\Delta\overline{yT} &= (p_1^*-p_0^*)\left\{(1-\alpha_1)\mathbb{E}\left[Y(1)|J=c\right]-\alpha_0\mathbb{E}\left[Y(0)|J=c\right]\right\}\\
&= (1-\alpha_1)\left\{\frac{\mathbb{E}\left[Y(1)-Y(0)|J=c\right]}{1-\alpha_0-\alpha_1}(p_1-p_0)\right\}+(p_1-p_0)\mathbb{E}\left[Y(0)|J=c\right]
\end{aligned}
$$

Recall that the analogous expression in the homogeneous treatment effect case is

$$
\begin{aligned}
\Delta\overline{yT} &= (1-\alpha_1)\mathcal{W}(p_1-p_0)+\mu_{10}^*\\
&= (1-\alpha_1)\left(\frac{\beta}{1-\alpha_0-\alpha_1}\right)(p_1-p_0)+(p_1-\alpha_0)m_{11}^*-(p_0-\alpha_0)m_{10}^*
\end{aligned}
$$

while the expression for the difference of variances is

$$
\Delta\overline{y^2}=\beta\mathcal{W}(p_1-p_0)+2\mathcal{W}\mu_{10}^*
$$

From above we see that the analogue of μ_{10}^* in the heterogeneous treatment effects setting is $(p_1-p_0)E\left[Y(0)|J=c\right]$ and since the LATE is $\mathbb{E}\left[Y(1)-Y(0)|J=c\right]$, the analogue of $\mathcal{W}$ is

$$
\frac{\mathbb{E}\left[Y(1)-Y(0)|J=c\right]}{1-\alpha_0-\alpha_1}
$$

so *if* we could establish that

$$\Delta\overline{y^2} = \left(\frac{p_1 - p_0}{1 - \alpha_0 - \alpha_1}\right) \mathbb{E}\left[Y(1) - Y(0)|J = c\right] \cdot \mathbb{E}\left[Y(1) + Y(0)|J = c\right]$$

in the heterogeneous treatment effects case, the proof of Theorem 2 would go through immediately. Now, if we assume an exclusion restriction on the *second* moment of y an argument almost identical to the standard LATE derivation gives

$$\Delta\overline{y^2} = \frac{\mathbb{E}\left[Y^2(1) - Y^2(0)|J = c\right]}{p_1^* - p_0^*} = \left(\frac{p_1 - p_0}{1 - \alpha_0 - \alpha_1}\right) \mathbb{E}\left[Y^2(1) - Y^2(0)|J = c\right]$$

so we see that the necessary and sufficient condition for our proof to go through is

$$\mathbb{E}\left[Y^2(1) - Y^2(0)|J = c\right] = \mathbb{E}\left[Y(1) - Y(0)|J = c\right] \cdot \mathbb{E}\left[Y(1) + Y(0)|J = c\right]$$

Rearranging, this in turn is equivalent to

$$\mathrm{Var}\left[Y(1)|J = c\right] = \mathrm{Var}\left[Y(0)|J = c\right]$$

9.3 Partial Identification Under Independence Assumption

Suppose we only make the LATE independence assumption $Y(T^*, z) = Y(T^*)$ rather than the conditional independence assumption $P(Y < \tau|T^*, z_k) = P(Y < \tau|T^*, z_\ell)$. Then we still obtain

$$\begin{aligned}
\mathbb{P}\left(Y|T^* = 1, z = 1\right) &= \left(\frac{p_0^*}{p_1^*}\right) \mathbb{P}\left[Y(1)|J = a\right] + \left(\frac{p_1^* - p_0^*}{p_1^*}\right) \mathbb{P}\left[Y(1)|J = c\right] \\
\mathbb{P}\left(Y|T^* = 0, z = 0\right) &= \left(\frac{p_1^* - p_0^*}{1 - p_0^*}\right) \mathbb{P}\left[Y(0)|J = c\right] + \left(\frac{1 - p_1^*}{1 - p_0^*}\right) \mathbb{P}\left[Y(0)|J = n\right] \\
\mathbb{P}\left(Y|T^* = 1, z = 0\right) &= \mathbb{P}\left[Y(1)|J = a\right] \\
\mathbb{P}\left(Y|T^* = 0, z = 1\right) &= \mathbb{P}\left[Y(0)|J = n\right]
\end{aligned}$$

From above, we also know that

$$\begin{aligned}
P(Y|T^* = 0, z_k) &= P(Y|T = 0, z_k) + \left(\frac{\alpha_1 p_k}{1 - p_k - \alpha_1}\right) \left[P(Y|T = 0, z_k) - P(Y|T = 1, z_k)\right] \\
P(Y|T^* = 1, z_k) &= P(Y|T = 1, z_k) + \left(\frac{\alpha_0(1 - p_k)}{p_k - \alpha_0}\right) \left[P(Y|T = 1, z_k) - P(Y|T = 0, z_k)\right]
\end{aligned}$$

The notation is getting a bit unwieldy so let $\pi^*_{tk}(y) = P(Y = y|T^* = t, z_k)$ and similarly define $\pi_{tk}(y) = P(Y = y|T = t, z_k)$. Using this new notation, we have

$$\begin{aligned}
(1 - p_k - \alpha_1)\pi^*_{0k}(y) &= (1 - p_k - \alpha_1)\pi_{0k}(y) + \alpha_1 p_k \left[\pi_{0k}(y) - \pi_{1k}(y)\right] \\
(p_k - \alpha_0)\pi^*_{1k}(y) &= (p_k - \alpha_0)\pi_{1k}(y) + \alpha_0(1 - p_k)\left[\pi_{1k}(y) - \pi_{0k}(y)\right]
\end{aligned}$$

Writing these out for all values of k,

$$\begin{aligned}
(p_1 - \alpha_0)\pi^*_{11}(y) &= (p_1 - \alpha_0)\pi_{11}(y) + \alpha_0(1 - p_1)\left[\pi_{11}(y) - \pi_{01}(y)\right] \\
(1 - p_0 - \alpha_1)\pi^*_{00}(y) &= (1 - p_0 - \alpha_1)\pi_{00}(y) + \alpha_1 p_0 \left[\pi_{00}(y) - \pi_{10}(y)\right] \\
(p_0 - \alpha_0)\pi^*_{10}(y) &= (p_0 - \alpha_0)\pi_{10}(y) + \alpha_0(1 - p_0)\left[\pi_{10}(y) - \pi_{00}(y)\right] \\
(1 - p_1 - \alpha_1)\pi^*_{01}(y) &= (1 - p_1 - \alpha_1)\pi_{01}(y) + \alpha_1 p_1 \left[\pi_{01}(y) - \pi_{11}(y)\right]
\end{aligned}$$

Similarly, using the fact that $p^*_k = (p_k - \alpha_0)/(1 - \alpha_0 - \alpha_1)$,

$$\begin{aligned}
\pi^*_{11}(y) &= \left(\frac{p_0 - \alpha_0}{p_1 - \alpha_0}\right) P\left[Y(1)|J = a\right] + \left(\frac{p_1 - p_0}{p_1 - \alpha_0}\right) P\left[Y(1)|J = c\right] \\
\pi^*_{00}(y) &= \left(\frac{p_1 - p_0}{1 - p_0 - \alpha_1}\right) P\left[Y(0)|J = c\right] + \left(\frac{1 - p_1 - \alpha_1}{1 - p_0 - \alpha_1}\right) P\left[Y(0)|J = n\right] \\
\pi^*_{10}(y) &= P\left[Y(1)|J = a\right] \\
\pi^*_{01}(y) &= P\left[Y(0)|J = n\right]
\end{aligned}$$

or equivalently,

$$\begin{aligned}
(p_1 - \alpha_0)\pi^*_{11}(y) &= (p_0 - \alpha_0)\, P\left[Y(1)|J = a\right] + (p_1 - p_0)\, P\left[Y(1)|J = c\right] \\
(1 - p_0 - \alpha_1)\pi^*_{00}(y) &= (p_1 - p_0)\, P\left[Y(0)|J = c\right] + (1 - p_1 - \alpha_1)\, P\left[Y(0)|J = n\right] \\
(p_0 - \alpha_0)\pi^*_{10}(y) &= (p_0 - \alpha_0) P\left[Y(1)|J = a\right] \\
(1 - p_1 - \alpha_1)\pi^*_{01}(y) &= (1 - p_1 - \alpha_1) P\left[Y(0)|J = n\right]
\end{aligned}$$

Equating,

$$\begin{aligned}
(p_0 - \alpha_0)\, P\left[Y(1)|J = a\right] + (p_1 - p_0)\, P\left[Y(1)|J = c\right] &= (p_1 - \alpha_0)\pi_{11}(y) + \alpha_0(1 - p_1)\left[\pi_{11}(y) - \pi_{01}(y)\right] \\
(p_1 - p_0)\, P\left[Y(0)|J = c\right] + (1 - p_1 - \alpha_1)\, P\left[Y(0)|J = n\right] &= (1 - p_0 - \alpha_1)\pi_{00}(y) + \alpha_1 p_0 \left[\pi_{00}(y) - \pi_{10}(y)\right] \\
(p_0 - \alpha_0) P\left[Y(1)|J = a\right] &= (p_0 - \alpha_0)\pi_{10}(y) + \alpha_0(1 - p_0)\left[\pi_{10}(y) - \pi_{00}(y)\right] \\
(1 - p_1 - \alpha_1) P\left[Y(0)|J = n\right] &= (1 - p_1 - \alpha_1)\pi_{01}(y) + \alpha_1 p_1 \left[\pi_{01}(y) - \pi_{11}(y)\right]
\end{aligned}$$

and substituting the third and fourth equalities into the first and second we obtain

$$\begin{aligned}
(p_0-\alpha_0)\pi_{10}(y)+\alpha_0(1-p_0)\left[\pi_{10}(y)-\pi_{00}(y)\right]+(p_1-p_0)\,P\left[Y(1)|J=c\right] &= (p_1-\alpha_0)\pi_{11}(y)+\alpha_0(1-p_1)\left[\pi_{11}(y)-\pi_{01}(y)\right]\\
(p_1-p_0)\,P\left[Y(0)|J=c\right]+(1-p_1-\alpha_1)\pi_{01}(y)+\alpha_1 p_1\left[\pi_{01}(y)-\pi_{11}(y)\right] &= (1-p_0-\alpha_1)\pi_{00}(y)+\alpha_1 p_0\left[\pi_{00}(y)-\pi_{10}(y)\right]\\
(p_0-\alpha_0)P\left[Y(1)|J=a\right] &= (p_0-\alpha_0)\pi_{10}(y)+\alpha_0(1-p_0)\left[\pi_{10}(y)-\pi_{00}(y)\right]\\
(1-p_1-\alpha_1)P\left[Y(0)|J=n\right] &= (1-p_1-\alpha_1)\pi_{01}(y)+\alpha_1 p_1\left[\pi_{01}(y)-\pi_{11}(y)\right]
\end{aligned}$$

Simplifying and re-arranging,

$$\begin{aligned}
P\left[Y(1)=y|J=c\right] &= \left[\frac{p_1\pi_{11}(y)-p_0\pi_{10}(y)}{p_1-p_0}\right]-\alpha_0\left[\frac{p_1\pi_{11}(y)-p_0\pi_{10}(y)+(1-p_1)\pi_{01}(y)-(1-p_0)\pi_{00}(y)}{p_1-p_0}\right]\\
P\left[Y(0)=y|J=c\right] &= \left[\frac{(1-p_0)\pi_{00}(y)-(1-p_1)\pi_{01}(y)}{p_1-p_0}\right]-\alpha_1\left[\frac{(1-p_0)\pi_{00}(y)-(1-p_1)\pi_{01}(y)+p_0\pi_{10}(y)-p_1\pi_{11}(y)}{p_1-p_0}\right]\\
P\left[Y(1)=y|J=a\right] &= \pi_{10}(y)+\left[\frac{\alpha_0(1-p_0)}{p_0-\alpha_0}\right]\left[\pi_{10}(y)-\pi_{00}(y)\right]\\
P\left[Y(0)=y|J=n\right] &= \pi_{01}(y)+\left[\frac{\alpha_1 p_1}{1-p_1-\alpha_1}\right]\left[\pi_{01}(y)-\pi_{11}(y)\right]
\end{aligned}$$

Notice that the first two equations can be simplified as follows

$$\begin{aligned}
P\left[Y(1)=y|J=c\right] &= \left[\frac{P(Y=y,T=1|z=1)-P(Y,T=1|z=0)}{p_1-p_0}\right]-\alpha_0\left[\frac{P(Y=y|z=1)-P(Y=y|z=0)}{p_1-p_0}\right]\\
P\left[Y(0)=y|J=c\right] &= \left[\frac{P(Y=y,T=0|z=1)-P(Y=y,T=0|z=1)}{p_1-p_0}\right]-\alpha_1\left[\frac{P(Y=y|z=0)-P(Y=y|z=1)}{p_1-p_0}\right]
\end{aligned}$$

Now, since probabilities must be between zero and one, we obtain the bounds

$$\begin{aligned}
0 &\leq \left[\frac{P(Y=y,T=1|z=1)-P(Y=y,T=1|z=0)}{p_1-p_0}\right]-\alpha_0\left[\frac{P(Y=y|z=1)-P(Y=y|z=0)}{p_1-p_0}\right]\leq 1\\
0 &\leq \left[\frac{P(Y=y,T=0|z=1)-P(Y=y,T=0|z=1)}{p_1-p_0}\right]-\alpha_1\left[\frac{P(Y=y|z=0)-P(Y=y|z=1)}{p_1-p_0}\right]\leq 1
\end{aligned}$$

which we abbreviate

$$\begin{aligned}
0 &\leq \left[\frac{\Delta P(Y=y,T=1)}{p_1-p_0}\right]-\alpha_0\left[\frac{\Delta P(Y=y)}{p_1-p_0}\right]\leq 1\\
0 &\leq \alpha_1\left[\frac{\Delta P(Y=y)}{p_1-p_0}\right]-\left[\frac{\Delta P(Y=y,T=0)}{p_1-p_0}\right]\leq 1
\end{aligned}$$

where

$$\begin{aligned}
\Delta P(Y=y) &= P(Y=y|z=1)-P(Y=y|z=0)\\
\Delta P(Y=y,T=t) &= P(Y=y,T=t|z=1)-P(Y=y,T=t|z=0).
\end{aligned}$$

To manipulate these bounds, we need to know the sign of $R = \Delta P(Y = y)/(p_1 - p_0)$. Presumably this will be positive for most values of y, but it could be negative.

Case I: R is positive.

$$\frac{\Delta P(Y=y,T=1)-(p_1-p_0)}{\Delta P(Y=y)} \leq \alpha_0 \leq \frac{\Delta P(Y=y,T=1)}{\Delta P(Y=y)}$$
$$\frac{\Delta P(Y=y,T=0)}{\Delta P(Y=y)} \leq \alpha_1 \leq \frac{\Delta P(Y=y,T=0)+(p_1-p_0)}{\Delta P(Y=y)}$$

Case II: R is negative.

$$\frac{\Delta P(Y=y,T=1)}{\Delta P(Y=y)} \leq \alpha_0 \leq \frac{\Delta P(Y=y,T=1)-(p_1-p_0)}{\Delta P(Y=y)}$$
$$\frac{\Delta P(Y=y,T=0)+(p_1-p_0)}{\Delta P(Y=y)} \leq \alpha_1 \leq \frac{\Delta P(Y=y,T=0)}{\Delta P(Y=y)}$$

Note that we *two-sided* bounds for the misclassification probabilities. These may be trivial in some cases, but I don't think it's obvious that they always will be.

Do these bounds have anything to do with the testability of the LATE assumptions? That is, do we get a lower bound for measurement error *precisely when* we would otherwise violate a testable LATE assumption?

Note that we also obtain bounds from the potential outcome distributions of always-takers and never-takers, namely

$$0 \leq \pi_{10}(y) + \left[\frac{\alpha_0(1-p_0)}{p_0-\alpha_0}\right][\pi_{10}(y)-\pi_{00}(y)] \leq 1$$
$$0 \leq \pi_{01}(y) + \left[\frac{\alpha_1 p_1}{1-p_1-\alpha_1}\right][\pi_{01}(y)-\pi_{11}(y)] \leq 1$$

but these are redundant. From the assumption of non-differential measurement error, we already have

$$\pi^*_{0k} = \pi_{0k} + \left(\frac{\alpha_1 p_k}{1-p_k-\alpha_1}\right)(\pi_{0k}-\pi_{1k})$$
$$\pi^*_{1k} = \pi_{1k} + \left(\frac{\alpha_0(1-p_k)}{p_k-\alpha_0}\right)(\pi_{1k}-\pi_{0k})$$

for all k as given at the beginning of this section. These expressions imply

$$0 \leq \pi_{0k} + \left(\frac{\alpha_1 p_k}{1 - p_k - \alpha_1}\right)(\pi_{0k} - \pi_{1k}) \leq 1$$
$$0 \leq \pi_{1k} + \left(\frac{\alpha_0(1 - p_k)}{p_k - \alpha_0}\right)(\pi_{1k} - \pi_{0k}) \leq 1$$

Re-arranging, we have

$$0 \leq (1 - p_k)\pi_{0k} - \alpha_1\pi_{0k} + \alpha_1 p_k (\pi_{0k} - \pi_{1k}) \leq 1 - p_k - \alpha_1$$
$$0 \leq p_k\pi_{1k} - \alpha_0\pi_{1k} + \alpha_0(1 - p_k)(\pi_{1k} - \pi_{0k}) \leq p_k - \alpha_0$$

and thus

$$0 \leq (1 - p_k)\pi_{0k} - \alpha_1 [(1 - p_k)\pi_{0k} + p_k\pi_{1k}] \leq 1 - p_k - \alpha_1$$
$$0 \leq p_k\pi_{1k} - \alpha_0 [p_k\pi_{1k} + (1 - p_k)\pi_{0k}] \leq p_k - \alpha_0$$

Now consider the first inequality. Re-arranging the right-hand side we obtain

$$\alpha_1 \leq \frac{(1 - p_k)(1 - \pi_{0k})}{1 - [(1 - p_k)\pi_{0k} + p_k\pi_{1k}]} = (1 - p_k)\left[\frac{P(Y = 0|T = 0, z = k)}{P(Y = 0|z = k)}\right]$$

and re-arranging the left-hand side we find

$$\alpha_1 \leq \frac{(1 - p_k)\pi_{0k}}{(1 - p_k)\pi_{0k} + p_k\pi_{1k}} = (1 - p_k)\left[\frac{P(Y = 1|T = 0, z = k)}{P(Y = 1|z = k)}\right]$$

For the second inequality, the left-hand side gives

$$\alpha_0 \leq \frac{p_k\pi_{1k}}{p_k\pi_{1k} + (1 - p_k)\pi_{0k}} = p_k\left[\frac{P(Y = 1|T = 1, z = k)}{P(Y = 1|z_k)}\right]$$

while the right-hand side gives

$$\alpha_0 \leq \frac{p_k(1 - \pi_{1k})}{1 - [p_k\pi_{1k} + (1 - p_k)\pi_{0k}]} = p_k\left[\frac{P(Y = 0|T = 1, z = k)}{P(Y = 0|z = k)}\right]$$

These are analogous to our CDF bounds from above although they may not be tighter than the bounds

$$\alpha_0 \leq p_k, \quad \alpha_1 \leq (1 - p_k)$$

because we cannot argue, as we did above, about a limit in which the ratio of CDFs approaches one. As before, however, we can take the tightest bound over $k = 0, 1$.

9.4 Bounding the LATE

Even if we didn't know anything about α_0 and α_1 beyond the fact that they are probabilities, it looks like we could still bound the LATE. I think we can do this without using the independence of the instrument, that is only using the mean exclusion restriction. Write out the LATE expressions with the α_0 and α_1 in them and them just plug in zero and one. Could then tighten the bounds by imposing additional assumptions to get bounds for α_0 and α_1, from weakest to strongest. If you have an independent instrument, you also get bounds for the outcome distributions. Need to think some more about this...

9.5 Stochastic Dominance Conditions

What if we imposed a stochastic ordering, e.g. $Y(1) > Y(0)$ for compliers? Presumably this would give joint bounds for α_0 and α_1 from the LATE expressions from above. Alternatively, perhaps one would choose to impose an ordering on the $Y(0)$ distributions for compliers versus never-takers or the $Y(1)$ distributions for the compliers versus always-takers. This might be interesting in situations where one is concerned that the assumption we need for identification does not in fact hold and should give additional bounds.

10 Outline For New Draft

1. Introduction / Literature Review
   (a) Why is this an important question?
      - Treatments of interest in economics usually endogenous and often binary.
      - Randomized encouragement designs are common in applied work.
      - Treatment status is often self-reported.
      - This problem is much more challenging that people realize.

   (b) Why are we different from Ura?
      - Main difference is that we, in line with the existing literature, study the case of non-differential measurement error. This allows us to obtain point identification under certain assumptions.
      - In contrast, Ura considers arbitray forms of mis-classification but as a consequence presents only partial identification results.

- Second, while we do provide results for LATE in Section blah, we mainly focus on additively separable model in which heterogeneity is captured by observed covariates while Ura considers only a LATE model. (And also doesn't allow for covariates.)

2. Mahajan/Lewbell-style Assumptions

   (a) Setup and Assumptions:

      - Homogenous treatment effect model (additively separable)
      - Conditional mean version of non-differential measurement error assumption.
      - Conditional mean independence for IV.

   (b) Show that the model is not identified, regardless of (discrete) support of IV.

   (c) Derive sharp bounds for α_0, α_1 and treatment effect.

   (d) Show that second and third-moment independence for IV identifies this model? Maybe this isn't interesting in and of itself?

3. Independence Assumption

   (a) Motivation

      - Showed above that stronger assumptions are needed for identification, but the additional moment restrictions seem a bit artificial.
      - When instruments are derived from economic theory that yields conditional mean independence only, we wouldn't want to use them.
      - They would make sense, however, in an an RCT or natural experiment.
      - The whole point in these settings is *not* to rely on functional form assumptions. It would be strange to say that z is an instrument for y but not $\log y$.
      - This points towards an *independence* assumption for the instrument.
      - Can make a similar argument for measurement error: seems strange to assume that T is non-differential for y but not $\log y$.

   (b) Sharp Bounds for α_0 and α_1 without valid instrument

      - Assume "independence" version of non-differential measurement error.
      - Derive CDF bounds.

   (c) Conditional Independence for Instrument

      - Exactly what assumptions do we need here?
      - Characteristic functions.

- Identification conditions?
- Overidentifying restrictions? Test model?

4. LATE Model

   (a) Introduction

   - Most of the existing mis-classification literature focuses on a homogeneous treatment effects model.
   - What if we don't have an additively separable model?
   - These results complement Ura because we work under the assumption of non-differential measurement error while he asks what can be learned when one is unwilling to make any assumptions about the form of the mis-classification.

   (b) Mahajan/Lewbel Setup

   - Presumably the partial identification results go through for a LATE.
   - The second and third moment conditions would require restrictions on form of heterogeneity. These would seem to be satisfied by a generalized Roy model.

   (c) Independence Assumptions

   - Presumably the CDF bounds go through as before but need to state exact form of independence assumption in terms of potential outcomes.
   - Kitagawa-style independence assumption for IV: $Y(T^*, z) = Y(T^*)$. This gives bounds for all quantile treatment effects.
   - Stochastic dominance conditions?

5. Estimation / Inference

6. Simulation Study

7. Empirical Examples

   - Try to look at a number of examples under different assumptions to illustrate both point and partial identification results. Don't forget about Oreopoulous: the sample size is so huge that inference isn't a major concern.

11 Weak Identification

11.1 Moment Equations

First we write the moment equations in a more familiar GMM-style form.

First Moment Condition: This is simply the IV moment condition:

$$\text{Cov}(y, z)/\text{Cov}(T, z) = \beta/(1 - \alpha_0 - \alpha_1)$$

Rearranging gives a more "canonical" GMM form:

$$\text{Cov}(y, z) - \left(\frac{\beta}{1 - \alpha_0 - \alpha_1}\right) \text{Cov}(T, z) = 0$$

Second Moment Condition: The equations used to identify $(\alpha_0 - \alpha_1)$ in the paper are

$$\begin{aligned}
\mu^*_{k\ell} &= (p_k - \alpha_0)m^*_{1k} - (p_\ell - \alpha_0)m^*_{1\ell} \\
\Delta\overline{y^2} &= \beta\mathcal{W}(p_k - p_\ell) + 2\mathcal{W}\mu^*_{k\ell} \\
\Delta\overline{yT} &= (1 - \alpha_1)\mathcal{W}(p_k - p_\ell) + \mu^*_{k\ell}
\end{aligned}$$

Re-arranging the third equation, $\mu^*_{k\ell} = \Delta\overline{yT} - (1 - \alpha_1)\mathcal{W}(p_k - p_\ell)$. Substituting into the second equation,

$$\begin{aligned}
\Delta\overline{y^2} &= \mathcal{W}\left[\beta(p_k - p_\ell) + 2\mu^*_{k\ell}\right] \\
&= \mathcal{W}\left\{\beta(p_k - p_\ell) + 2\left[\Delta\overline{yT} - (1 - \alpha_1)\mathcal{W}(p_k - p_\ell)\right]\right\} \\
&= \mathcal{W}\left\{(p_k - p_\ell)\left[\beta - 2(1 - \alpha_1)\mathcal{W}\right] + 2\Delta\overline{yT}\right\}
\end{aligned}$$

Now, substituting $\mathcal{W} = \beta/(1 - \alpha_0 - \alpha_1)$

$$\begin{aligned}
\Delta\overline{y^2} &= \frac{\beta}{1 - \alpha_0 - \alpha_1}\left\{\beta(p_k - p_\ell)\left[1 - \frac{2(1 - \alpha_1)}{1 - \alpha_0 - \alpha_1}\right] + 2\Delta\overline{yT}\right\} \\
&= \frac{\beta}{1 - \alpha_0 - \alpha_1}\left\{2\Delta\overline{yT} - \beta(p_k - p_\ell)\left(\frac{1 + \alpha_0 - \alpha_1}{1 - \alpha_0 - \alpha_1}\right)\right\}
\end{aligned}$$

We now write this in a more standard form. Let w be any random variable. Then,

$$\begin{aligned}
\text{Cov}(w, z) &= E(wz) - E(w)E(z) = \left[1 \times E(w|z = 1)q + 0 \times E(w|z = 0)(1 - q)\right] - E(w)q \\
&= qE(w|z = 1) - qE(w) = qE(w|z = 1) - q\left[E(w|z = 1)q + E(w|z = 0)(1 - q)\right] \\
&= q\left[E(w|z = 1)(1 - q) + E(w|z = 0)(1 - q)\right] \\
&= q(1 - q)\left[E(w|z = 1) - E(w|z = 0)\right]
\end{aligned}$$

Using this fact, we can express the quantities that appear in the second moment equality in terms of covariances as follows

$$\Delta\overline{y^2} = \frac{\text{Cov}(y^2, z)}{q(1-q)}, \quad \Delta\overline{yT} = \frac{\text{Cov}(yT, z)}{q(1-q)}, \quad (p_k - p_\ell) = \frac{\text{Cov}(T, z)}{q(1-q)}$$

leading to

$$\frac{\text{Cov}(y^2, z)}{q(1-q)} = \frac{\beta}{1-\alpha_0-\alpha_1}\left\{\frac{2\text{Cov}(yT, z)}{q(1-q)} - \beta\frac{\text{Cov}(T, z)}{q(1-q)}\left(\frac{1+\alpha_0-\alpha_1}{1-\alpha_0-\alpha_1}\right)\right\}$$

Or, multiplying through by $q(1-q)$ and re-arranging,

$$\text{Cov}(y^2, z) - \frac{\beta}{1-\alpha_0-\alpha_1}\left\{2\text{Cov}(yT, z) - \beta\text{Cov}(T, z)\left(\frac{1+\alpha_0-\alpha_1}{1-\alpha_0-\alpha_1}\right)\right\} = 0$$

Third Moment Condition: The third and final set of moment conditions is

$$\begin{aligned}\Delta\overline{y^3} &= \beta^2[\mathcal{W}(p_k - p_\ell)] + 3\beta[\mathcal{W}\mu^*_{k\ell}] + 3\mathcal{W}\lambda^*_{k\ell}\\ \Delta\overline{y^2T} &= \beta(1-\alpha_1)\mathcal{W}(p_k - p_\ell) + 2(1-\alpha_1)\mathcal{W}\mu^*_{k\ell} + \lambda^*_{k\ell}\end{aligned}$$

To put this into a more familiar format, we first eliminate $\mu^*_{k\ell}$ using

$$\mu^*_{k\ell} = \Delta\overline{yT} - (1-\alpha_1)\mathcal{W}(p_k - p_\ell)$$

from the derivation of the second moment equation from above, yielding

$$\begin{aligned}\Delta\overline{y^3} &= \beta^2[\mathcal{W}(p_k - p_\ell)] + 3\beta\mathcal{W}\left[\Delta\overline{yT} - (1-\alpha_1)\mathcal{W}(p_k - p_\ell)\right] + 3\mathcal{W}\lambda^*_{k\ell}\\ \Delta\overline{y^2T} &= \beta(1-\alpha_1)\mathcal{W}(p_k - p_\ell) + 2(1-\alpha_1)\mathcal{W}\left[\Delta\overline{yT} - (1-\alpha_1)\mathcal{W}(p_k - p_\ell)\right] + \lambda^*_{k\ell}\end{aligned}$$

Re-arranging and factoring the first equation gives

$$\Delta\overline{y^3} = \mathcal{W}\left(p_k - p_\ell\right)\left\{\beta^2 + \frac{3\beta\Delta\overline{yT}}{p_k - p_\ell} - 3\beta\mathcal{W}(1-\alpha_1) + \frac{3\lambda^*_{k\ell}}{p_k - p_\ell}\right\}$$

Now, by re-arranging the second equation we find that

$$\begin{aligned}\lambda^*_{k\ell} &= \Delta\overline{y^2T} - \beta(1-\alpha_1)\mathcal{W}(p_k - p_\ell) - 2(1-\alpha_1)\mathcal{W}\left[\Delta\overline{yT} - (1-\alpha_1)\mathcal{W}(p_k - p_\ell)\right]\\ &= \Delta\overline{y^2T} - \beta(1-\alpha_1)\mathcal{W}(p_k - p_\ell) - 2(1-\alpha_1)\mathcal{W}\Delta\overline{yT} + 2(1-\alpha_1)^2\mathcal{W}^2(p_k - p_\ell)\end{aligned}$$

and thus

$$\frac{3\lambda^*_{k\ell}}{p_k - p_\ell} = 3\left(\frac{\Delta\overline{y^2 T}}{p_k - p_\ell}\right) - 3\beta(1-\alpha_1)\mathcal{W} - 6(1-\alpha_1)\mathcal{W}\left(\frac{\Delta\overline{yT}}{p_k - p_\ell}\right) + 6(1-\alpha_1)^2\mathcal{W}^2$$

so that

$$\begin{aligned}
\frac{\Delta\overline{y^3}}{\mathcal{W}(p_k - p_\ell)} &= \left\{\beta^2 - 6\beta\mathcal{W}(1-\alpha_1) + 6\mathcal{W}^2(1-\alpha_1)^2 + \left(\frac{\Delta\overline{yT}}{p_k - p_\ell}\right)[3\beta - 6\mathcal{W}(1-\alpha_1)] + 3\left(\frac{\Delta\overline{y^2 T}}{p_k - p_\ell}\right)\right\} \\
&= \left\{\beta^2\left[1 - \frac{6(1-\alpha_1)}{1-\alpha_0-\alpha_1} + \frac{6(1-\alpha_1)^2}{(1-\alpha_0-\alpha_1)^2}\right] + 3\beta\left[1 - \frac{2(1-\alpha_1)}{1-\alpha_0-\alpha_1}\right]\left(\frac{\Delta\overline{yT}}{p_k - p_\ell}\right) + 3\left(\frac{\Delta\overline{y^2 T}}{p_k - p_\ell}\right)\right\}
\end{aligned}$$

Simplifying, we find that

$$\left[1 - \frac{2(1-\alpha_1)}{1-\alpha_0-\alpha_1}\right] = \frac{(1-\alpha_0-\alpha_1) - 2(1-\alpha_1)}{1-\alpha_0-\alpha_1} = \frac{1-\alpha_0-\alpha_1-2+2\alpha_1}{1-\alpha_0-\alpha_1} = \frac{\alpha_1-\alpha_0-1}{1-\alpha_0-\alpha_1}$$

and

$$\begin{aligned}
\left[1 - \frac{6(1-\alpha_1)}{1-\alpha_0-\alpha_1} + \frac{6(1-\alpha_1)^2}{(1-\alpha_0-\alpha_1)^2}\right] &= 1 - \left[\frac{6(1-\alpha_1)}{1-\alpha_0-\alpha_1}\right]\left[1 - \frac{1-\alpha_1}{1-\alpha_0-\alpha_1}\right] \\
&= 1 - \left[\frac{6(1-\alpha_1)}{1-\alpha_0-\alpha_1}\right]\left[\frac{(1-\alpha_0-\alpha_1)-(1-\alpha_1)}{1-\alpha_0-\alpha_1}\right] \\
&= 1 + \frac{6\alpha_0(1-\alpha_1)}{(1-\alpha_0-\alpha_1)^2}
\end{aligned}$$

so that

$$\frac{\Delta\overline{y^3}}{\mathcal{W}(p_k - p_\ell)} = \left\{\beta^2\left[1 + \frac{6\alpha_0(1-\alpha_1)}{(1-\alpha_0-\alpha_1)^2}\right] - 3\beta\left[\frac{1-(\alpha_1-\alpha_0)}{1-\alpha_0-\alpha_1}\right]\left(\frac{\Delta\overline{yT}}{p_k - p_\ell}\right) + 3\left(\frac{\Delta\overline{y^2 T}}{p_k - p_\ell}\right)\right\}$$

Therefore, re-arranging and multiplying through by $q(1-q)$,

$$\mathrm{Cov}(z, y^3) = \left(\frac{\beta}{1-\alpha_0-\alpha_1}\right)\left\{\beta^2\left[1 + \frac{6\alpha_0(1-\alpha_1)}{(1-\alpha_0-\alpha_1)^2}\right]\mathrm{Cov}(z,T) - 3\beta\left[\frac{1-(\alpha_1-\alpha_0)}{1-\alpha_0-\alpha_1}\right]\mathrm{Cov}(z,yT) + 3\mathrm{Cov}(z,y^2T)\right\}$$

Full Set of Moment Conditions

$$\mathrm{Cov}(y,z) - \left(\frac{\beta}{1-\alpha_0-\alpha_1}\right)\mathrm{Cov}(T,z) = 0$$

$$\mathrm{Cov}(y^2,z) - \frac{\beta}{1-\alpha_0-\alpha_1}\left\{2\mathrm{Cov}(yT,z) - \beta\mathrm{Cov}(T,z)\left(\frac{1+\alpha_0-\alpha_1}{1-\alpha_0-\alpha_1}\right)\right\} = 0$$

$$\mathrm{Cov}(y^3,z) - \left(\frac{\beta}{1-\alpha_0-\alpha_1}\right)\left\{\beta^2\left[1 + \frac{6\alpha_0(1-\alpha_1)}{(1-\alpha_0-\alpha_1)^2}\right]\mathrm{Cov}(T,z) - 3\beta\left[\frac{1-(\alpha_1-\alpha_0)}{1-\alpha_0-\alpha_1}\right]\mathrm{Cov}(yT,z) + 3\mathrm{Cov}(y^2T,z)\right\} = 0$$

11.2 Simple Special Case: $\alpha_0 = 0$

Suppose that α_0. Then the model is identified using the first and second moment equalities, which simplify to

$$\begin{aligned}
\text{Cov}(y,z) - \left(\frac{\beta}{1-\alpha_1}\right)\text{Cov}(T,z) &= 0\\
\text{Cov}(y^2,z) - \left(\frac{\beta}{1-\alpha_1}\right)\left[2\text{Cov}(yT,z) - \beta\text{Cov}(T,z)\right] &= 0
\end{aligned}$$

In this simple special case, it is easy to solve for β by substituting the first moment condition into the second:

$$\beta = \frac{2\text{Cov}(yT,z)}{\text{Cov}(T,z)} - \frac{\text{Cov}(y^2,z)}{\text{Cov}(y,z)}$$

I checked this equation in our simulation experiment and it is indeed correct

Notice that if $\beta \approx 0$ then both $\text{Cov}(y^2,z)$ and $\text{Cov}(y,z)$ are close to zero so their ratio becomes extremely noisy.

Standard GMM form: To express this system in the standard GMM form, we need to agument these moment equalities with expressions for the means of z, y, y^2, T, and yT as follows. Let $\mathbf{w}_i = (y_i, z_i, T_i)'$, $\theta = (\beta, \alpha_1)'$ and $\gamma = (q, p, \mu, s, r)'$ where

$$\begin{aligned}
q &= \mathbb{E}[z]\\
p &= \mathbb{E}[T]\\
\mu &= \mathbb{E}[y]\\
s &= \mathbb{E}\left[y^2\right]\\
r &= \mathbb{E}[yT].
\end{aligned}$$

We can express our problem in terms of two blocks of moment conditions, namely

$$f(\mathbf{w};\theta,\gamma) = \begin{bmatrix} g(\mathbf{w};\theta,\gamma) \\ h(\mathbf{w};\gamma) \end{bmatrix}$$

where

$$g(\mathbf{w};\theta,\gamma) = \begin{bmatrix} (zy - q\mu) - \left(\frac{\beta}{1-\alpha_1}\right)(zT - qp) \\ (zy^2 - qs) - 2\left(\frac{\beta}{1-\alpha_1}\right)(zyT - qr) + \left(\frac{\beta^2}{1-\alpha_1}\right)(zT - qp) \end{bmatrix}$$

and

$$h(\mathbf{w};\gamma) = \begin{bmatrix} z-q \\ T-p \\ y-\mu \\ y^2-s \\ yT-r \end{bmatrix}$$

We can view this as a two-step or "plug-in" GMM estimation problem where $\widehat{\gamma}$ solves the sample moment condition

$$\frac{1}{n}\sum_{i=1}^{n} h(\mathbf{w}_i;\gamma) = 0$$

and $\widehat{\theta}$ solves

$$\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i;\theta,\widehat{\gamma}) = 0.$$

Unfortunately, in our example the first-step estimation affects the asymptotic variance of the second since an inconsistent estimator of γ yields an inconsistent estimator of θ.[2] This means that we will have to proceed "the hard way."

Under standard regularity conditions, a GMM estimator based on the sample analogue $f_n(\theta,\gamma)$ of $\mathbb{E}[f(\mathbf{w};\theta,\gamma)] = 0$ using a weighting matrix $\widehat{W} \to_p W$ converges in distribution to

$$-(F'WF)^{-1}F'WM, \quad M \sim N(0,\Omega)$$

where $\sqrt{n}f_n(\theta_0,\gamma_0) \to_d M$ and $F = \mathbb{E}[\nabla'_\theta f(\mathbf{w};\theta_0,\gamma_0), \nabla'_\gamma f(\mathbf{w};\theta_0,\gamma_0)]$. The present example, however, is just-identified which means that F is square and hence

$$-(F'WF)^{-1}F'W = F^{-1}W^{-1}(F')^{-1}F'W = -F^{-1}$$

Now, given the special structure of our example,

$$F = \begin{bmatrix} \mathbb{E}\left\{\nabla'_\theta g(\mathbf{w};\theta_0,\gamma_0)\right\} & \mathbb{E}\left\{\nabla'_\gamma g(\mathbf{w};\theta_0,\gamma_0)\right\} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \equiv \begin{bmatrix} G_\theta & G_\gamma \\ \mathbf{0} & -\mathbf{I} \end{bmatrix}$$

becuase h does not involve θ and $\nabla'_\gamma h(\mathbf{w},\gamma) = -\mathbf{I}$. Inverting, we have

$$-F^{-1} = \begin{bmatrix} -G_\theta & -G_\gamma \\ \mathbf{0} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} -G_\theta^{-1} & -G_\theta^{-1}G_\gamma \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

[2]See Newey & McFadden (1994), Section 6.

We see from this expression that if G_γ were zero, the first step-estimation would not affect the limit distribution of $\widehat{\theta}$. Differentiating,

$$\begin{bmatrix} \nabla g_\beta & \nabla g_{\alpha_1} \end{bmatrix} = \begin{bmatrix} -\left(\dfrac{zT - qp}{1-\alpha_1}\right) & -\left\{\dfrac{\beta(zT-qp)}{(1-\alpha_1)^2}\right\} \\ 2\left\{\dfrac{\beta(zT-qp) - (zyT - qr)}{1-\alpha_1}\right\} & \dfrac{\beta^2(zT-qp) - 2\beta(zyT-qr)}{(1-\alpha_1)^2} \end{bmatrix}$$

and thus, taking expectations,

$$G_\theta = \begin{bmatrix} \dfrac{-\mathrm{Cov}(z,T)}{1-\alpha_1} & \dfrac{-\beta\mathrm{Cov}(z,T)}{(1-\alpha_1)^2} \\ 2\left\{\dfrac{\beta\mathrm{Cov}(z,T) - \mathrm{Cov}(yT,z)}{1-\alpha_1}\right\} & \dfrac{\beta^2\mathrm{Cov}(z,T) - 2\beta\mathrm{Cov}(yT,z)}{(1-\alpha_1)^2} \end{bmatrix}$$

Now, for G_γ we have

$$\begin{aligned} G_\gamma &= \mathbb{E}\begin{bmatrix} \nabla_q g & \nabla_p \mu & \nabla_\mu g & \nabla_s g & \nabla_r g \end{bmatrix} \\ &= \begin{bmatrix} \left(\dfrac{p\beta}{1-\alpha_1} - \mu\right) & \left(\dfrac{q\beta}{1-\alpha_1}\right) & -q & 0 & 0 \\ \left(\dfrac{\beta}{1-\alpha_1}\right)(2r - \beta p) - s & \dfrac{-q\beta^2}{1-\alpha_1} & 0 & -q & \dfrac{2\beta q}{1-\alpha_1} \end{bmatrix} \end{aligned}$$

The next step is to invert G_θ. First we calculate the determinant. For the purposes of this calculation, use the shorthand $C = \mathrm{Cov}(z,T)$ and $D = \mathrm{Cov}(yT,z)$. We have:

$$\begin{aligned} |G_\theta| &= \left[\frac{-C}{1-\alpha_1}\right]\left[\frac{\beta^2 C - 2\beta D}{(1-\alpha_1)^2}\right] - \left[\frac{-\beta C}{(1-\alpha_1)^2}\right]\left[\frac{2\beta C - 2D}{1-\alpha_1}\right] \\ &= \left(\frac{1}{1-\alpha_1}\right)^3 \left[2\beta CD - \beta^2 C^2 + 2\beta^2 C^2 - 2\beta CD\right] \\ &= \frac{\beta^2 \mathrm{Cov}(z,T)^2}{(1-\alpha_1)^3} \end{aligned}$$

Thus,

$$G_\theta^{-1} = \frac{(1-\alpha_1)^3}{\beta^2 \text{Cov}(z,T)^2} \left[\begin{array}{cc} \dfrac{\beta^2 \text{Cov}(z,T) - 2\beta \text{Cov}(yT,z)}{(1-\alpha_1)^2} & \dfrac{\beta \text{Cov}(z,T)}{(1-\alpha_1)^2} \\ \\ -2\left\{ \dfrac{\beta \text{Cov}(z,T) - \text{Cov}(yT,z)}{1-\alpha_1} \right\} & \dfrac{-\text{Cov}(z,T)}{1-\alpha_1} \end{array} \right]$$

$$= \left[\begin{array}{cc} \left\{ \dfrac{1-\alpha_1}{\text{Cov}(z,T)} \right\} \left\{ 1 - \dfrac{2\text{Cov}(yT,z)}{\beta \text{Cov}(z,T)} \right\} & \dfrac{1-\alpha_1}{\beta \text{Cov}(z,T)} \\ \\ \dfrac{2(1-\alpha_1)^2}{\beta \text{Cov}(z,T)} \left\{ \dfrac{\text{Cov}(yT,z)}{\beta \text{Cov}(z,T)} - 1 \right\} & \dfrac{-(1-\alpha_1)^2}{\beta^2 \text{Cov}(z,T)} \end{array} \right]$$

The next step is to calculate Ω:

$$\Omega = \lim_{n\to\infty} \text{Var}\left[\sqrt{n} f_n(\theta_0, \gamma_0)\right] = \lim_{n\to\infty} \text{Var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n f(\mathbf{w}_i; \theta_0, \gamma_0)\right]$$

If $\mathbf{w}_i$ is an iid sequence of RVs, then

$$\Omega = \lim_{n\to\infty} \frac{1}{n} \text{Var}\left[\sum_{i=1}^n f(\mathbf{w}_i; \theta_0, \gamma_0)\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \text{Var}\left[f(\mathbf{w}_i; \theta_0, \gamma_0)\right] = \text{Var}\left[f(\mathbf{w}_i; \theta_0, \gamma_0)\right]$$

And assuming that our model is correctly specified, so that $\mathbb{E}\left[f(\mathbf{w}_i; \theta_0, \gamma_0)\right] = 0$,

$$\begin{aligned} \text{Var}\left[f(\mathbf{w}_i; \theta_0, \gamma_0)\right] &= \mathbb{E}\left[\begin{array}{cc} g(\mathbf{w}_i; \theta_0, \gamma_0) g(\mathbf{w}_i; \theta_0, \gamma_0)' & g(\mathbf{w}_i; \theta_0, \gamma_0) h(\mathbf{w}_i; \theta_0, \gamma_0)' \\ h(\mathbf{w}_i; \theta_0, \gamma_0) g(\mathbf{w}_i; \theta_0, \gamma_0)' & h(\mathbf{w}_i; \theta_0, \gamma_0) h(\mathbf{w}_i; \theta_0, \gamma_0)' \end{array} \right] \\ &\equiv \left[\begin{array}{cc} \Omega_{gg} & \Omega_{gh} \\ \Omega_{gh} & \Omega_{hh} \end{array} \right] \end{aligned}$$

We now calculate each block.

I don't think this is actually going to give us anything interpretable. The expressions are quite involved and it seems unlikely that they'll cancel in a useful way. This doesn't matter for implementation, of course, since it's easy to calculate the estimate of $\widehat{\Omega}$ by plugging the GMM estimates into the sample moment conditions, taking the outer product, and averaging.

We are only interested in the asymptotic variance matrix V_θ of our parameters of interest

$\theta = (\beta, \alpha_1)$. We calculate this as follows:

$$\begin{aligned}
V_\theta &= \left[\begin{array}{cc} G_\theta^{-1} & G_\theta^{-1}G_\gamma \end{array}\right] \left[\begin{array}{cc} \Omega_{gg} & \Omega_{gh} \\ \Omega_{hg} & \Omega_{hh} \end{array}\right] \left[\begin{array}{c} \left(G_\theta^{-1}\right)' \\ \left(G_\theta^{-1}G_\gamma\right)' \end{array}\right] \\
&= \left[\begin{array}{cc} G_\theta^{-1}\left(\Omega_{gg} + G_\gamma \Omega_{gh}\right) & G_\theta^{-1}\left(\Omega_{gh} + G_\gamma \Omega_{hh}\right) \end{array}\right] \left[\begin{array}{c} \left(G_\theta^{-1}\right)' \\ G_\gamma' \left(G_\theta^{-1}\right)' \end{array}\right] \\
&= G_\theta^{-1}\left(\Omega_{gg} + G_\gamma \Omega_{hg}\right)\left(G_\theta^{-1}\right)' + G_\theta^{-1}\left(\Omega_{gh} + G_\gamma \Omega_{hh}\right) G_\gamma' \left(G_\theta^{-1}\right)' \\
&= G_\theta^{-1}\left(\Omega_{gg} + G_\gamma \Omega_{hg} + \Omega_{gh} G_\gamma' + G_\gamma \Omega_{hh} G_\gamma'\right)\left(G_\theta^{-1}\right)'
\end{aligned}$$

11.3 Easier(?) Derivation of Simple Special Case: $\alpha_0 = 0$

Recall that we could eliminate α_1 from the moment conditions, yielding,

$$\beta = \frac{2\text{Cov}(yT, z)}{\text{Cov}(T, z)} - \frac{\text{Cov}(y^2, z)}{\text{Cov}(y, z)}$$

We can treat this as our g block of moment conditions with a parameter vector θ that is simply β. This gives

$$g(\mathbf{w}; \beta, \gamma) = \left[2\left(\frac{zTy - qr}{zT - qp}\right) - \frac{zy^2 - qs}{zy - q\mu} - \beta\right]$$

The h block of moment conditions is unchanged. Now, we have

$$F = \left[\begin{array}{cc} \mathbb{E}\left\{\nabla_\beta' g(\mathbf{w}; \beta_0, \gamma_0)\right\} & \mathbb{E}\left\{\nabla_\gamma' g(\mathbf{w}; \beta_0, \gamma_0)\right\} \\ \mathbf{0} & -\mathbf{I} \end{array}\right] \equiv \left[\begin{array}{cc} G_\beta & G_\gamma \\ \mathbf{0} & -\mathbf{I} \end{array}\right] = \left[\begin{array}{cc} -1 & G_\gamma \\ \mathbf{0} & -\mathbf{I} \end{array}\right]$$

since the derivative of g with respect to β is -1 and that of h with respect to γ is $-\mathbf{I}$. Inverting,

$$-F^{-1} = \left[\begin{array}{cc} -G_\beta & -G_\gamma \\ \mathbf{0} & \mathbf{I} \end{array}\right]^{-1} = \left[\begin{array}{cc} -G_\beta^{-1} & -G_\beta^{-1}G_\gamma \\ \mathbf{0} & \mathbf{I} \end{array}\right] = \left[\begin{array}{cc} 1 & G_\gamma \\ \mathbf{0} & \mathbf{I} \end{array}\right]$$

We calculate G_γ as follows:

$$G_\gamma = \mathbb{E}\left[\begin{array}{ccccc} \nabla_q g & \nabla_p g & \nabla_\mu g & \nabla_s g & \nabla_r g \end{array}\right]$$

$$\begin{aligned}
\nabla_q g(\mathbf{w}; \beta_0, \gamma_0) &= \nabla_q \left[2\left(\frac{zTy - qr}{zT - qp}\right) - \frac{zy^2 - qs}{zy - q\mu} - \beta\right] \\
&= 2\left[\frac{-r(zT - qp) + p(zTy - qr)}{(zT - qp)^2}\right] - \frac{-s(zy - q\mu) + \mu(zy^2 - qs)}{(zy - q\mu)^2} \\
&= \frac{2zT(py - r)}{(zT - qp)^2} - \frac{zy(\mu y - s)}{(zy - q\mu)^2}
\end{aligned}$$

$$\nabla_p g(\mathbf{w}; \beta_0, \gamma_0) = \frac{2q(zTy - qr)}{(zT - qp)^2}$$

$$\nabla_\mu g(\mathbf{w}; \beta_0, \gamma_0) = \frac{-q(zy^2 - qs)}{(zy - q\mu)^2}$$

$$\nabla_s g(\mathbf{w}; \beta_0, \gamma_0) = \frac{q}{zy - q\mu}$$

$$\nabla_r g(\mathbf{w}; \beta_0, \gamma_0) = \frac{-2q}{zT - qp}$$

The next step is to calculate Ω:

$$\Omega = \text{Var}\left[f(\mathbf{w}_i; \theta_0, \gamma_0)\right] = \mathbb{E}\left[\begin{array}{cc} g(\mathbf{w}_i; \theta_0, \gamma_0) g(\mathbf{w}_i; \theta_0, \gamma_0)' & g(\mathbf{w}_i; \theta_0, \gamma_0) h(\mathbf{w}_i; \theta_0, \gamma_0)' \\ h(\mathbf{w}_i; \theta_0, \gamma_0) g(\mathbf{w}_i; \theta_0, \gamma_0)' & h(\mathbf{w}_i; \theta_0, \gamma_0) h(\mathbf{w}_i; \theta_0, \gamma_0)' \end{array}\right]$$

$$\equiv \left[\begin{array}{cc} \Omega_{gg} & \Omega_{gh} \\ \Omega_{gh} & \Omega_{hh} \end{array}\right]$$

We now calculate each block.

Still need to do this!

11.4 Two-Step Inference Idea

If β is small, then confidence intervals based on the GMM limit distribution from above will perform badly. But even in this case, inference for the *identified set* $[\beta_{RF}, \beta_{IV}]$ should still be well-behaved, so long as the instrument is strong. The idea of this section is to explore a two-step procedure that chooses between reporting the GMM confidence interval or inference for the identified set based on a pre-test of β_{RF}. Presumably conducting valid inference based on such a procedure will require a Bonferroni correction. The first step, however, is to determine the joint limiting behavior of the reduced form, IV, and GMM estimators.

Reduced Form Estimator The reduced form is given by

$$\begin{aligned} y &= \gamma_0 + \gamma_1 z + \eta \\ \gamma_0 &= \mathbb{E}[y|z=0] \\ \gamma_1 &= \mathbb{E}[y|z=0] - \mathbb{E}[y|z=1] \end{aligned}$$

Now, we need to write η in terms of the "primitives" of our model. The first stage and main equation are

$$\begin{aligned} y &= c + \beta T^* + \varepsilon \\ T^* &= \pi_0 + \pi_1 z + v \\ \pi_1 &= p_1^* - p_0^* \end{aligned}$$

which implies

$$y = (c + \beta\pi_0) + (\beta\pi_1)z + (\varepsilon + \beta v)$$

so that

$$\begin{aligned} \gamma_1 &= \beta(p_1^* - p_0^*) \\ \eta &= \varepsilon + \beta v \end{aligned}$$

Now, define $W = (\mathbf{1}, \mathbf{z})$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$. Then the reduced form estimator is

$$\widehat{\boldsymbol{\gamma}} = (W'W)^{-1} W'\mathbf{y} = (W'W)^{-1} W'(W\boldsymbol{\gamma} + \boldsymbol{\eta}) = \boldsymbol{\gamma} + (W'W)^{-1} W'\boldsymbol{\eta}$$

and hence

$$\sqrt{n}\,(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \left(\frac{W'W}{n}\right)^{-1} \frac{W'\boldsymbol{\eta}}{\sqrt{n}}$$

Now,

$$\left(\frac{W'W}{n}\right)^{-1} = \begin{bmatrix} 1 & \bar{\mathbf{z}} \\ \bar{\mathbf{z}} & \mathbf{z}'\mathbf{z}/n \end{bmatrix}^{-1} \to_p \begin{bmatrix} 1 & q \\ q & q \end{bmatrix}^{-1} = \frac{1}{q(1-q)} \begin{bmatrix} q & -q \\ -q & 1 \end{bmatrix}$$

and by the Central Limit Theorem,

$$\frac{W'\boldsymbol{\eta}}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i\left[\begin{array}{c}1\\ z_i\end{array}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\varepsilon_i+\beta v_i)\left[\begin{array}{c}1\\ z_i\end{array}\right] \rightarrow_d N\left(\mathbf{0},\Sigma\right)$$

where

$$\begin{aligned}\Sigma = \mathbb{E}\left[\begin{array}{cc}\eta_i^2 & z_i\eta_i^2\\ z_i\eta_i^2 & z_i^2\eta_i^2\end{array}\right] &= \mathbb{E}\left[\begin{array}{cc}(\varepsilon_i+\beta v_i)^2 & z_i\left(\varepsilon_i+\beta v_i\right)^2\\ z_i\left(\varepsilon_i+\beta v_i\right)^2 & z_i^2\left(\varepsilon_i+\beta v_i\right)^2\end{array}\right]\\ &= \mathbb{E}\left[\begin{array}{cc}\varepsilon_i^2+2\beta\varepsilon_i v_i+\beta^2 v_i^2 & z_i^2\varepsilon_i+2\beta z_i\varepsilon_i v_i+\beta^2 z_i v_i^2\\ z_i^2\varepsilon_i+2\beta z_i\varepsilon_i v_i+\beta^2 z_i v_i^2 & z_i^2\varepsilon_i^2+2\beta z_i^2\varepsilon_i v_i+\beta^2 z_i^2 v_i^2\end{array}\right]\end{aligned}$$

The next step is to work out the joint distribution of v and the other primitives of our model:

$$\begin{aligned}&T^*=0, z=0 \implies 0=\pi_0+v \implies v=-\pi_0=-p_0^*\\ &T^*=1, z=0 \implies 1=\pi_0+v \implies v=1-\pi_0=1-p_0^*\\ &T^*=0, z=1 \implies 0=\pi_0+\pi_1+v \implies v=-(\pi_0+\pi_1)=-p_1^*\\ &T^*=1, z=1 \implies 1=\pi_0+\pi_1+v \implies v=1-(\pi_0+\pi_1)=1-p_1^*\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}(T^*=0,z=0) &= \mathbb{P}(v=-p_0^*) = (1-p_0^*)(1-q)\\ \mathbb{P}(T^*=1,z=0) &= \mathbb{P}(v=1-p_0^*) = p_0^*(1-q)\\ \mathbb{P}(T^*=0,z=1) &= \mathbb{P}(v=-p_1^*) = (1-p_1^*)q\\ \mathbb{P}(T^*=1,z=1) &= \mathbb{P}(v=1-p_1^*) = p_1^* q\end{aligned}$$

Notice that, as must be true *by construction*, v is mean zero:

$$\begin{aligned}\mathbb{E}[v] &= -p_0^*(1-p_0^*)(1-q)+(1-p_0^*)p_0^*(1-q)-p_1^*(1-p_1^*)q+(1-p_1^*)p_1^*q\\ &= (1-q)(1-p_0^*)(p_0^*-p_0^*)+q(1-p_1^*)(p_1^*-p_1^*)=0\end{aligned}$$

and uncorrelated with v:

$$\begin{aligned}\mathbb{E}[zv] &= q\mathbb{E}[v|z=1] = q\left\{p_1^*\mathbb{E}[v|T^*=1,z=1]+(1-p_1^*)\mathbb{E}[v|T^*=0,z=1]\right\}\\ &= q\left[p_1^*(1-p_1^*)-p_1^*(1-p_1^*)\right]=0\end{aligned}$$

Now, to calculate Σ, we need $\mathbb{E}[v^2], \mathbb{E}[\varepsilon v], \mathbb{E}[z\varepsilon v], \mathbb{E}[z^2\varepsilon v], \mathbb{E}[zv^2]$, and $\mathbb{E}[z^2v^2]$:

$$\begin{aligned}
\mathbb{E}[v^2] &= p_0^{*2}(1-p_0^*)(1-q) + (1-p_0^*)^2 p_0^*(1-q) + p_1^{*2}(1-p_1^*)q + (1-p_1^*)^2 p_1^* q \\
&= (1-q)p_0^*(1-p_0^*)\left[p_0^* + (1-p_0^*)\right] + qp_1^*(1-p_1^*)\left[p_1^* + (1-p_1^*)\right] \\
&= (1-q)p_0^*(1-p_0^*) + qp_1^*(1-p_1^*)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\varepsilon v] &= \mathbb{E}\left[v\mathbb{E}\left(\varepsilon|v\right)\right] \\
&= -p_0^*\mathbb{P}(v=-p_0^*)\mathbb{E}(\varepsilon|v=-p_0^*) - p_1^*\mathbb{P}(v=-p_1^*)\mathbb{E}(\varepsilon|v=-p_1^*) \\
&\qquad + (1-p_0^*)\mathbb{P}(v=1-p_0^*)\mathbb{E}(\varepsilon|v=1-p_0^*) + (1-p_1^*)\mathbb{P}(v=1-p_1^*)\mathbb{E}(\varepsilon|v=1-p_1^*) \\
&= -p_0^*(1-p_0^*)(1-q)\mathbb{E}[\varepsilon|T^*=0,z=0] - p_1^*(1-p_1^*)q\mathbb{E}[\varepsilon|T^*=0,z=1] \\
&\qquad + (1-p_0^*)p_0^*(1-q)\mathbb{E}[\varepsilon|T^*=1,z=0] + (1-p_1^*)p_1^*q\mathbb{E}[\varepsilon|T^*=1,z=1] \\
&= -p_0^*(1-p_0^*)(1-q)(m_{00}^*-c) - p_1^*(1-p_1^*)q(m_{01}^*-c) \\
&\qquad + (1-p_0^*)p_0^*(1-q)(m_{10}^*-c) + (1-p_1^*)p_1^*q(m_{11}^*-c) \\
&= (1-q)p_0^*(1-p_0^*)(m_{10}^*-m_{00}^*) + qp_1^*(1-p_1^*)(m_{11}^*-m_{01}^*)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[z\varepsilon v] = \mathbb{E}[z^2\varepsilon v] = \mathbb{E}\left[z^2\mathbb{E}\left[\varepsilon v|z\right]\right] &= q\mathbb{E}\left[\varepsilon v|z=1\right] = q\mathbb{E}_{T^*|z=1}\left[\mathbb{E}\left[\varepsilon v|T^*=1,z=1\right]\right] \\
&= q\left\{p_1^*\mathbb{E}\left[\varepsilon v|\,T^*=1,z=1\right] + (1-p_1^*)\mathbb{E}[\varepsilon v|T^*=0,z=1]\right\} \\
&= q\left\{p_1^*(1-p_1^*)\mathbb{E}\left[\varepsilon|\,T^*=1,z=1\right] - p_1^*(1-p_1^*)\mathbb{E}[\varepsilon|T^*=0,z=1]\right\} \\
&= qp_1^*(1-p_1^*)\left[m_{11}^*-m_{01}^*\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[zv^2] = \mathbb{E}[z^2v^2] = \mathbb{E}\left[z^2\mathbb{E}\left[v^2|z\right]\right] &= q\mathbb{E}[v^2|z=1] = q\mathbb{E}_{T^*|z=1}\left[\mathbb{E}\left[v^2|T^*,z=1\right]\right] \\
&= q\left\{p_1^*\mathbb{E}[v^2|T^*=1,z=1] + (1-p_1^*)\mathbb{E}\left[v^2|T^*=0,z=1\right]\right\} \\
&= q\left\{p_1^*(1-p_1^*)^2 + p_1^{*2}(1-p_1^*)\right\} \\
&= qp_1^*(1-p_1^*)\left[(1-p_1^*) + p_1^*\right] \\
&= qp_1^*(1-p_1^*)
\end{aligned}$$

Using these calculations, we find the elements of Σ as follows:

$$\begin{aligned}
\mathbb{E}[\eta^2] &= \mathbb{E}[\varepsilon^2 + 2\beta\varepsilon v + \beta^2 v^2] \\
&= \sigma_\varepsilon^2 + 2\beta\left[(1-q)p_0^*(1-p_0^*)(m_{10}^* - m_{00}^*) + qp_1^*(1-p_1^*)(m_{11}^* - m_{01}^*)\right] \\
&\quad + \beta^2\left[(1-q)p_0^*(1-p_0^*) + qp_1^*(1-p_1^*)\right] \\
&= \sigma_\varepsilon^2 + (1-q)p_0^*(1-p_0^*)\beta\left[\beta + 2\left(m_{10}^* - m_{00}^*\right)\right] + qp_1^*(1-p_1^*)\beta\left[\beta + 2\left(m_{11}^* - m_{01}^*\right)\right] \\
\mathbb{E}[z^2\eta^2] = \mathbb{E}[z\eta^2] &= \mathbb{E}[z\varepsilon^2 + 2\beta z\varepsilon v + \beta^2 z v^2] = \sigma_\varepsilon^2 + 2\beta\mathbb{E}[z\varepsilon v] + \beta^2\mathbb{E}[zv^2] \\
&= q\sigma_\varepsilon^2 + 2\beta q p_1^*(1-p_1^*)\left(m_{11}^* - m_{01}^*\right) + \beta^2 q p_1^*(1-p_1^*) \\
&= q\left\{\sigma_\varepsilon^2 + p_1^*(1-p_1^*)\beta\left[\beta + 2\left(m_{11}^* - m_{01}^*\right)\right]\right\}
\end{aligned}$$

Now, using the fact that $\mathbb{E}[z^2\eta] = \mathbb{E}[z\eta]$, the asymptotic variance of the reduced form estimator can be written as:

$$\begin{aligned}
\text{AVAR}\left[\sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right)\right] &= \frac{1}{q^2(1-q)^2}\begin{bmatrix} q & -q \\ -q & 1 \end{bmatrix}\begin{bmatrix} \mathbb{E}(\eta^2) & \mathbb{E}(z\eta^2) \\ \mathbb{E}(z\eta^2) & \mathbb{E}(z\eta^2) \end{bmatrix}\begin{bmatrix} q & -q \\ -q & 1 \end{bmatrix} \\
&= \frac{1}{q^2(1-q)^2}\begin{bmatrix} q & -q \\ -q & 1 \end{bmatrix}\begin{bmatrix} q\left\{\mathbb{E}(\eta^2) - \mathbb{E}(z\eta^2)\right\} & \mathbb{E}(z\eta^2) - q\mathbb{E}(\eta^2) \\ 0 & (1-q)\mathbb{E}(z\eta^2) \end{bmatrix} \\
&= \frac{1}{q^2(1-q)^2}\begin{bmatrix} q^2\left\{\mathbb{E}(\eta^2) - \mathbb{E}(z\eta^2)\right\} & -q^2\left\{\mathbb{E}(\eta^2) - \mathbb{E}(z\eta^2)\right\} \\ -q^2\left\{\mathbb{E}(\eta^2) - \mathbb{E}(z\eta^2)\right\} & q^2\mathbb{E}(\eta^2) + (1-2q)\mathbb{E}(z\eta^2) \end{bmatrix}
\end{aligned}$$

Now, we are only interested in the reduced form slope coefficient γ_1. The asymptotic variance of its OLS estimator is:

$$\begin{aligned}
\text{AVAR}\left[\sqrt{n}(\widehat{\gamma}_1 - \gamma_1)\right] &= \frac{1}{q^2(1-q)^2}\left[q^2\mathbb{E}(\eta^2) + (1-2q)\mathbb{E}(z\eta^2)\right] \\
&= \frac{\mathbb{E}(\eta^2)}{(1-q)^2} + \left[\frac{1-2q}{q^2(1-q)^2}\right]\mathbb{E}(z\eta^2)
\end{aligned}$$

In general, this will lead to quite a complicated expression. In our simulation design, however,

$$\begin{aligned}
q &= 1/2 \\
p_0^*(1-p_0^*) &= \delta(1-\delta) \\
p_1^*(1-p_1^*) &= (1-\delta)\delta \\
(m_{10}^* - m_{00}^*) &= (m_{11}^* - m_{01}^*) > 0
\end{aligned}$$

leading to the following simplifications:

$$\begin{aligned}
\text{AVAR}\left[\sqrt{n}(\widehat{\gamma}_1 - \gamma_1)\right] &= 4\mathbb{E}(\eta^2) \\
&= 4\left\{\sigma_\varepsilon^2 + 1/2 \times \delta(1-\delta)\beta\left[\beta + 2\left(m_{10}^* - m_{00}^*\right)\right] + 1/2 \times \delta(1-\delta)\beta\left[\beta + 2\left(m_{11}^* - m_{01}^*\right)\right]\right\} \\
&= 4\left\{\sigma_\varepsilon^2 + \delta(1-\delta)\beta\left[\beta + 2\left(m_{11}^* - m_{01}^*\right)\right]\right\}
\end{aligned}$$

Notice that, since $(m_{11}^* - m_{01}^*)$ is positive, as in the simulation design, the asymptotic variance is *smallest* when $\beta = 0$ so that there is no treatment effect.

Robust Standard Errors When implementing the reduced form estimator in practice we base our inference on sample residuals:

$$\widehat{\eta}_i = y_i - \widehat{\gamma}_0 - \widehat{\gamma}_1 z_i$$

As we saw above, however, the errors η are heteroskedastic: $\mathbb{E}(z\eta^2) \neq \mathbb{E}(z)\mathbb{E}(\eta^2)$ etc. For this reason, we must use robust standard errors:

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \widehat{\eta}_i^2 & z_i\widehat{\eta}_i^2 \\ z_i\widehat{\eta}_i^2 & z_i^2\widehat{\eta}_i^2 \end{bmatrix}$$

leading to

$$\widehat{\text{AVAR}}\left[\sqrt{n}(\widehat{\gamma}_1 - \gamma_1)\right] = \frac{\widehat{\mathbb{E}}(\eta^2)}{(1-\widehat{q})^2} + \left[\frac{1-2\widehat{q}}{\widehat{q}^2(1-\widehat{q})^2}\right]\widehat{\mathbb{E}}(z\eta^2)$$

Note that $\widehat{q}$ is fixed and equal to 1/2 in our simulation design, so this becomes $4\widehat{\sigma}_\eta^2$.

Values of m_{tk}^* in the Simulation To calculate Σ in our simulation design, we'll need to know the values of m_{tk}^* in the threshold-crossing model with bivariate normal errors:

$$\begin{bmatrix} \varepsilon \\ \xi \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

$$\begin{aligned}
T^* &= \mathbf{1}\left\{\kappa_0 + \kappa_1 z + \xi > 0\right\} \\
\kappa_0 &= \Phi^{-1}(\delta) \\
\kappa_1 &= \Phi^{-1}(1-\delta) - \Phi^{-1}(\delta)
\end{aligned}$$

$$
\begin{aligned}
m^*_{00} - c &= \mathbb{E}[\varepsilon | T^* = 0, z = 0] = \mathbb{E}[\varepsilon | \xi \leq -\kappa_0] \\
m^*_{01} - c &= \mathbb{E}[\varepsilon | T^* = 0, z = 1] = \mathbb{E}[\varepsilon | \xi \leq -(\kappa_0 + \kappa_1)] \\
m^*_{10} - c &= \mathbb{E}[\varepsilon | T^* = 1, z = 0] = [\varepsilon | \xi > -\kappa_0] \\
m^*_{11} - c &= \mathbb{E}[\varepsilon | T^* = 1, z = 1] = \mathbb{E}[\varepsilon | \xi > -(\kappa_0 + \kappa_1)]
\end{aligned}
$$

$$
\xi | \varepsilon \sim N(\rho\varepsilon, 1 - \rho^2)
$$

$$
\begin{aligned}
\mathbb{P}(\varepsilon \leq x | \xi \leq a) &= \frac{\mathbb{P}(\varepsilon \leq x, \xi \leq a)}{\mathbb{P}(\xi \leq a)} = \frac{\int_{-\infty}^{x} \int_{-\infty}^{a} f(\xi|\varepsilon) f(\varepsilon) d\xi \, d\varepsilon}{\Phi(a)} \\
&= \frac{\int_{-\infty}^{x} F_{\xi|\varepsilon}(a) f(\varepsilon) d\varepsilon}{\Phi(a)} = \frac{\displaystyle\int_{-\infty}^{x} \varphi(\varepsilon) \Phi\left(\frac{a - \rho\varepsilon}{\sqrt{1-\rho^2}}\right) d\varepsilon}{\Phi(a)}
\end{aligned}
$$

$$
f(\varepsilon | \xi \leq a) = \frac{\varphi(\varepsilon)}{\Phi(a)} \Phi\left(\frac{a - \rho\varepsilon}{\sqrt{1-\rho^2}}\right)
$$

$$
\mathbb{E}\left[\varepsilon | \xi \leq a\right] = \frac{1}{\Phi(a)} \int_{-\infty}^{\infty} \varepsilon \varphi(\varepsilon) \Phi\left(\frac{a - \rho\varepsilon}{\sqrt{1-\rho^2}}\right) d\varepsilon
$$

$$
\begin{aligned}
\mathbb{P}(\varepsilon \leq x | \xi > a) &= \frac{\mathbb{P}(\varepsilon \leq x, \xi > a)}{\mathbb{P}(\xi > a)} = \frac{\int_{-\infty}^{x} \int_{a}^{\infty} f(\xi|\varepsilon) f(\varepsilon) d\xi \, d\varepsilon}{1 - \Phi(a)} \\
&= \frac{\int_{-\infty}^{x} [1 - F_{\xi|\varepsilon}(a)] f(\varepsilon) d\varepsilon}{1 - \Phi(a)} = \frac{\displaystyle\int_{-\infty}^{x} \varphi(\varepsilon) \left[1 - \Phi\left(\frac{a - \rho\varepsilon}{\sqrt{1-\rho^2}}\right)\right] d\varepsilon}{1 - \Phi(a)}
\end{aligned}
$$

$$
f(\varepsilon | \xi > a) = \frac{\varphi(\varepsilon)}{1 - \Phi(a)} \left[1 - \Phi\left(\frac{a - \rho\varepsilon}{\sqrt{1-\rho^2}}\right)\right]
$$

$$
\mathbb{E}\left[\varepsilon | \xi > a\right] = \frac{1}{1 - \Phi(a)} \int_{-\infty}^{\infty} \varepsilon \varphi(\varepsilon) \left[1 - \Phi\left(\frac{a - \rho\varepsilon}{\sqrt{1-\rho^2}}\right)\right] d\varepsilon
$$

I've checked all of these integrals in R and they're definitely correct.

IV Estimator First we calculate the probability limits of the IV estimators of β and c. Notice that *both* are inconsistent:

$$\begin{aligned}\widehat{\beta}_{IV} &= \left(\widetilde{Z}\widetilde{T}/n\right)^{-1}\left(\widetilde{Z}'\mathbf{y}/n\right) = \left(\widetilde{Z}\widetilde{T}/n\right)^{-1}\widetilde{Z}'\left(\widetilde{T}^*\widetilde{\beta}+\boldsymbol{\varepsilon}\right)/n \\ &= \left(\widetilde{Z}'\widetilde{T}/n\right)^{-1}\left(\left[\widetilde{Z}'\widetilde{T}^*/n\right]\widetilde{\beta}+\widetilde{Z}'\boldsymbol{\varepsilon}/n\right)\end{aligned}$$

$$\widetilde{Z}'\boldsymbol{\varepsilon}/n \to_p \mathbb{E}\begin{bmatrix}\varepsilon \\ z\varepsilon\end{bmatrix} = \mathbf{0}$$

$$\left(\widetilde{Z}'\widetilde{T}/n\right)^{-1} = \begin{bmatrix}1 & \bar{T} \\ \bar{z} & \mathbf{z}'\mathbf{T}/n\end{bmatrix} \to_p \begin{bmatrix}1 & p \\ q & p_1 q\end{bmatrix} = \frac{1}{q(p_1-p)}\begin{bmatrix}p_1 q & -p \\ -q & 1\end{bmatrix}$$

$$\widetilde{Z}'\widetilde{T}^*/n = \begin{bmatrix}1 & \bar{T}^* \\ \bar{z} & \mathbf{z}'\mathbf{T}^*/n\end{bmatrix} \to_p \begin{bmatrix}1 & p^* \\ q & p_1^* q\end{bmatrix}$$

$$\left(\widetilde{Z}'\widetilde{T}/n\right)^{-1}\left(\widetilde{Z}'\widetilde{T}^*/n\right) \to_p \frac{1}{q(p_1-p)}\begin{bmatrix}qp_1 & -p \\ -q & 1\end{bmatrix}\begin{bmatrix}1 & p^* \\ q & qp_1^*\end{bmatrix} = \begin{bmatrix}1 & (p_1p^*-p_1^*p)/(p_1-p) \\ 0 & (p_1^*-p^*)/(p_1-p)\end{bmatrix}$$

$$\frac{p_1^*-p^*}{p_1-p} = \frac{1}{p_1-p}\left(\frac{p_1-\alpha_0}{1-\alpha_0-\alpha_1} - \frac{p-\alpha_0}{1-\alpha_0-\alpha_1}\right) = \frac{1}{1-\alpha_0-\alpha_1}$$

$$\frac{p_1p^*-p_1^*p}{p_1-p} = \frac{1}{p_1-p}\left[\frac{p_1(p-\alpha_0)}{1-\alpha_0-\alpha_1} - \frac{p(p_1-\alpha_0)}{1-\alpha_0-\alpha_1}\right] = \frac{-\alpha_0}{1-\alpha_0-\alpha_1}$$

$$\boldsymbol{\beta}_{IV} \to_p \begin{bmatrix}1 & (p_1p^*-p_1^*p)/(p_1-p) \\ 0 & (p_1^*-p^*)/(p_1-p)\end{bmatrix}\begin{bmatrix}c \\ \beta\end{bmatrix} = \frac{1}{1-\alpha_0-\alpha_1}\begin{bmatrix}c-\alpha_0\beta \\ \beta\end{bmatrix} \equiv \begin{bmatrix}c_{IV} \\ \beta_{IV}\end{bmatrix}$$

Now that we have the probability limit of the IV estimator, we can work out the error term ω that corresponds to it as a function of the "primitives" of our model. We have:

$$\begin{aligned}
\zeta = y - (c_{IV} + \beta_{IV} T) &= (c + \beta T^* + \varepsilon) - \frac{1}{1-\alpha_0-\alpha_1}\left[(c - \alpha_0\beta) + \beta T\right] \\
&= \varepsilon + c - \left(\frac{c - \alpha_0 \beta}{1-\alpha_0-\alpha_1}\right) + \beta\left(T^* - \frac{T}{1-\alpha_0-\alpha_1}\right) \\
&= \varepsilon + \left[\frac{\alpha_0(\beta - c) - \alpha_1 c}{1-\alpha_0-\alpha_1}\right] + \beta\left[\frac{(T^* - T) - (\alpha_0+\alpha_1)T^*}{1-\alpha_0-\alpha_1}\right] \\
&= \varepsilon + \left[\frac{\alpha_0(\beta - c) - \alpha_1 c}{1-\alpha_0-\alpha_1}\right] + \beta\left[\frac{w - (\alpha_0+\alpha_1)(\pi_0 + \pi_1 z + v)}{1-\alpha_0-\alpha_1}\right]
\end{aligned}$$

This looks pretty complicated. I'm also not sure we need to work this out analytically. All we really need is to write out the joint limit distribution of the IV and RF...

Joint Distribution of IV and RF To carry out inference for the identified set, we need to work out the joint distribution of the reduced form and IV estimators:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{IV} &= \left(\widetilde{Z}'\widetilde{T}/n\right)^{-1}\left(\widetilde{Z}'\mathbf{y}/n\right) \\
\widehat{\boldsymbol{\gamma}} &= \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\left(\widetilde{Z}'\boldsymbol{y}/n\right)
\end{aligned}$$

The probability limits of each estimator define an associated error term, each of which depends in a complicated way on our "primitive" model parameters:

$$\begin{aligned}
y &= c_{IV} + \beta_{IV} T + \zeta \\
y &= \gamma_0 + \gamma_1 z + \eta
\end{aligned}$$

But because these two estimators depend only on observable quantities, we can use their residuals to work out the covariance matrix of (ζ, η). Accordingly, we proceed as follows:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{IV} &= \left(\widetilde{Z}'\widetilde{T}/n\right)^{-1}\left(\widetilde{Z}'\left[\widetilde{T}\boldsymbol{\beta}_{IV} + \boldsymbol{\zeta}\right]/n\right) = \boldsymbol{\beta}_{IV} + \left(\widetilde{Z}'\widetilde{T}/n\right)^{-1}\left(\widetilde{Z}'\boldsymbol{\zeta}/n\right) \\
\widehat{\boldsymbol{\gamma}} &= \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\left(\widetilde{Z}'\left[\widetilde{Z}\boldsymbol{\gamma} + \boldsymbol{\eta}\right]/n\right) = \boldsymbol{\gamma} + \left(\widetilde{Z}'\widetilde{Z}/n\right)^{-1}\left(\widetilde{Z}\boldsymbol{\eta}/n\right)
\end{aligned}$$

yielding

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_{IV}\right) = \left(\widetilde{Z}'\widetilde{T}/n\right)^{-1}\left(\widetilde{Z}'\boldsymbol{\zeta}/\sqrt{n}\right)$$
$$\sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = \left(\widetilde{Z}'\widetilde{Z}/n\right)^{-1}\left(\widetilde{Z}'\boldsymbol{\eta}/\sqrt{n}\right)$$

11.5 Other Stuff...

- Manski was interested in a heterogenous treatment effect model and whether we could bound ATE rather than $LATE$.
- Would it help in the performance of the GMM estimator if we enforced the bounds on α_0 and α_1?

Exogenous Covariates in a Linear Model: These should be very easy to handle because we can just stack the GMM moment conditions to include an IV estimator for the parameter on the exogenous covariates in the main equation. Recall that the usual IV estimator for these parameters is unaffected by measurement error. We should write this out since it's the case that many people will use in practice given the extreme sample size demands of fully non-parametric estimation!

More About Weak Identification: Sophocles pointed out in an email exchange that I had been assuming (incorrectly) that $\text{Cov}(y, z)$ is always well-behaved. This is not the case if z is a weak instrument. I don't think we can simply assume we have a strong instrument and consider the weak identification that arises from $\beta \approx 0$ in isolation. I think the two problems of $\beta \approx 0$ and weak z interact in an important way since, as we saw from above, the determinant $|G_\theta|$ that measures the strength of identification depends on the *product* of β and $Cov(z, T)$. I think it the correct interpretation of this is that the magnitude of β that gives strong identification should is always relative to the strength of z. If z is very strong, then β can be smaller without causing problems. But if z is weak then I think β needs to be really large to get strong identification. If I recall correctly, we uncovered something in our simulations that appears to agree with this intuition but I need to go back and check.

To see why $\text{Cov}(y, z)$ is badly behaved when z is weak, write out an explicit first-stage equation for our model as follows:

$$T^* = \pi_0 + \pi_1 z + v$$

where

$$\pi_0 = \mathbb{E}[T^*|z=0] = p_0^*$$
$$\pi_1 = \mathbb{E}[T^*=1|z=1] - \mathbb{E}[T^*|z=0] = p_1^* - p_0^*$$

and $\mathbb{E}[zv] = 0$ by construction. Now,

$$\begin{aligned}
\text{Cov}(z,y) &= \mathbb{E}(zy) - \mathbb{E}(z)\mathbb{E}(y) \\
&= \mathbb{E}\left[z\left(c + \beta T^* + \varepsilon\right)\right] - q\mathbb{E}\left(c + \beta T^* + \varepsilon\right) \\
&= \mathbb{E}\left[z\left\{c + \beta\left(\pi_0 + \pi_1 z + v\right) + \varepsilon\right\}\right] - q\mathbb{E}\left[c + \beta\left(\pi_0 + \pi_1 z + v\right) + \varepsilon\right] \\
&= q(c + \beta\pi_0) + \beta\pi_1\mathbb{E}(z^2) - q\left(c + \beta\pi_0 + \beta\pi_1\mathbb{E}[z]\right) \\
&= q(c + \beta\pi_0 + \beta\pi_1) - q\left(c + \beta\pi_0 + \beta\pi_1 q\right) \\
&= q(\beta\pi_1 - \beta\pi_1 q) = \beta\pi_1 q(1-q) \\
&= \beta(p_1^* - p_0^*)q(1-q)
\end{aligned}$$

Auxiliary Moment Inequalities Notice that if $\beta = 0$, then the preceding moment equalities do *not* identify α_1. However, we do have auxiliary moment *inequalities* that partially identify α_1 regardless of the value of β. The simplest of these comes from the relationship

$$p_k^* = \frac{p_k - \alpha_0}{1 - \alpha_0 - \alpha_1}, \quad 1 - p_k^* = \frac{1 - p_k - \alpha_1}{1 - \alpha_0 - \alpha_1}$$

where $p_k = P(T=1|z_k)$ and $p_k^* = P(T^*=1|z_k)$. (This follows from the Law of Total Probability and our assumption that the mis-classification probabilities rates depend only on T^*, not z.) Under our assumption that $\alpha_0 + \alpha_1 < 1$, we obtain $\alpha_0 < \min_k p_k$ and $\alpha_1 < \min_k(1 - p_k)$. If $\alpha_0 = 0$, as we assume in the present special case, then without any assumption on the true value of α_1 we have

$$0 \leq \alpha_1 < \min_k(1 - p_k) = 1 - \max_k p_k.$$

Is there some way to use these moment inequalities in estimation?

Under Normality In our simulation for the CDF bounds on α_0 and α_1, we found that the upper bounds were in fact equal to the true parameter values. This is very surprising and is very likely comes from the specific parametric model from which we simulated. This happens to have been a model with normally distributed errors. Can we say anything about such a model theoretically? Perhaps try to write down the likelihood function? This could also be a useful way to look at the weak identification problem.

12 April 2017 – New GMM Formulation

Consider the special case in which $\alpha_0 = 0$ so the model is identified from

$$\begin{aligned}\mathrm{Cov}(y,z) - \left(\frac{\beta}{1-\alpha_1}\right)\mathrm{Cov}(T,z) = 0\\ \mathrm{Cov}(y^2,z) - \left(\frac{\beta}{1-\alpha_1}\right)\left[2\mathrm{Cov}(yT,z) - \beta\mathrm{Cov}(T,z)\right] = 0\end{aligned}$$

Above we wrote this in a standard GMM form by adding auxiliary moment equations to identify $\mathbb{E}[y^2]$, $\mathbb{E}[yT]$, etc. But there's a simpler and more transparent way to do this. Under our assumptions and $\alpha_0 = 0$, some algebra shows that

Add the algebra from the whiteboard notes later

$$\mathbb{E}\left[y - \frac{\beta}{1-\alpha_1}T\right] = c$$

and that

$$\mathbb{E}\left[y^2 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha}T\right] = \sigma_{\varepsilon\varepsilon} + c^2$$

where c is the intercept from the regression model and $\sigma_{\varepsilon\varepsilon} = \mathrm{Var}(\varepsilon)$. Now, re-writing the first covariance equation using the linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[yz - \frac{\beta}{1-\alpha_1}Tz\right] - \mathbb{E}[z]\mathbb{E}\left[y - \frac{\beta}{1-\alpha}T\right] = 0\\ \mathbb{E}\left[yz - \frac{\beta}{1-\alpha_1}Tz\right] - \mathbb{E}[z]c = 0\\ \mathbb{E}\left[\left\{y - c - \frac{\beta}{1-\alpha_1}T\right\}z\right] = 0\end{aligned}$$

and proceeding similarly for the second covariance equation

$$\begin{aligned}\mathbb{E}\left[y^2z - \frac{\beta}{1-\alpha_1}2yTz + \frac{\beta^2}{1-\alpha_1}Tz\right] - \mathbb{E}[z]\mathbb{E}\left[y - \frac{\beta}{1-\alpha_1}yT + \frac{\beta^2}{1-\alpha_1}T\right] = 0\\ \mathbb{E}\left[y^2z - \frac{\beta}{1-\alpha_1}2yTz + \frac{\beta^2}{1-\alpha_1}Tz\right] - \mathbb{E}[z]\left(\sigma_{\varepsilon\varepsilon} + c^2\right) = 0\\ \mathbb{E}\left[\left\{y^2 - \sigma_{\varepsilon\varepsilon} - c^2 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha_1}T\right\}z\right] = 0\end{aligned}$$

Thus, we can express our estimator in terms of the following four moment equations:

$$\begin{aligned}
\mathbb{E}\left[y - c - \frac{\beta}{1-\alpha_1}T\right] &= 0\\
\mathbb{E}\left[y^2 - \sigma_{\varepsilon\varepsilon} - c^2 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha_1}T\right] &= 0\\
\mathbb{E}\left[\left\{y - c - \frac{\beta}{1-\alpha_1}T\right\}z\right] &= 0\\
\mathbb{E}\left[\left\{y^2 - \sigma_{\varepsilon\varepsilon} - c^2 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha_1}T\right\}z\right] &= 0
\end{aligned}$$

To simplify the notation, let $\boldsymbol{\theta} = (\alpha_1, \beta, c, \sigma_{\varepsilon\varepsilon})'$ and define

$$\begin{aligned}
u(\boldsymbol{\theta}) &= y - c - \frac{\beta}{1-\alpha_1}T\\
v(\boldsymbol{\theta}) &= y^2 - \sigma_{\varepsilon\varepsilon} - c^2 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha_1}
\end{aligned}$$

Then we can express the four moment equalities as

$$\mathbb{E}\left[g_1(\mathbf{x}, \boldsymbol{\theta})\right] = \mathbb{E}\begin{bmatrix} u(\boldsymbol{\theta}) \\ v(\boldsymbol{\theta}) \end{bmatrix} = \mathbf{0}, \quad \mathbb{E}\left[g_2(\mathbf{x}, \boldsymbol{\theta})\right] = \mathbb{E}\begin{bmatrix} u(\boldsymbol{\theta})z \\ v(\boldsymbol{\theta})z \end{bmatrix} = \mathbf{0}$$

We also have two moment inequalities, namely $\alpha_1 \leq 1 - p_1$ and $\alpha_1 \leq 1 - p_0$. After some algebra (see the whiteboard), we can show that these are equivalent to

$$\mathbb{E}\left[h(\mathbf{x}, \theta)\right] = \mathbb{E}\begin{bmatrix} (1-\alpha_1) - T(1-z)/(1-q) \\ (1-\alpha_1) - Tz/q \end{bmatrix} \geq \mathbf{0}$$

where $q = \mathbb{E}[z]$. We will *condition* on z, i.e. hold it fixed in repeated samples, so we will not add an extra moment condition for q. Instead we will simply substitute the sample analogue.

To formulate the GMM estimator with moment inequalities as in Moon and Schorfheide (2009), we introduce some further notation. Let $\boldsymbol{\lambda}$ denote the slack in $h(\mathbf{x}, \theta)$, so that

$$\mathbb{E}[h(\mathbf{x}, \theta)] = \boldsymbol{\lambda} \geq \mathbf{0} \iff \mathbb{E}[h(\mathbf{x}, \theta) - \boldsymbol{\lambda}] = \mathbf{0}$$

Further define

$$f(\mathbf{x}) = \begin{bmatrix} g(\mathbf{x}, \theta) \\ h(\mathbf{x}, \boldsymbol{\theta}) \end{bmatrix}, \quad \psi(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) = \begin{bmatrix} g(\mathbf{x}, \boldsymbol{\theta}) \\ h(\mathbf{x}, \boldsymbol{\theta}) - \boldsymbol{\lambda} \end{bmatrix}$$

and let $\Theta = \{\boldsymbol{\theta} \colon \alpha_1 \geq 0, \sigma_{\varepsilon\varepsilon} \geq 0\}$. Then, the GMM estimator based on our moment equalities and inequalities is

$$(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = \underset{\boldsymbol{\theta}\in\Theta, \boldsymbol{\lambda}\geq \mathbf{0}}{\arg\min} \quad Q_n(\boldsymbol{\theta}, \boldsymbol{\lambda})$$

where

$$Q_n(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{2}\bar{\psi}_n(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda})'\mathbf{W}_n\bar{\psi}_n(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda})$$

$$\bar{\psi}_n(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\psi(\mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\lambda})$$

and $\mathbf{W}_n$ is a weighting matrix that should be irrelevant in our case because we're just identified. We scale the criterion by 1/2 so that the derivative looks nice.

Is this still the case with the moment inequalities? Mechanically we have introduced two new parameters to match the two extra conditions.

At various points we will need the derivatives of the moment equations. Let

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbb{E}[\nabla_{\boldsymbol{\theta}'} g(\mathbf{x}, \boldsymbol{\theta})], \quad \mathbf{H}(\boldsymbol{\theta}) = \mathbb{E}[\nabla_{\boldsymbol{\theta}'} h(\mathbf{x}, \boldsymbol{\theta})]$$

and define

$$\begin{aligned}\boldsymbol{\Psi}(\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \mathbb{E}\left[\begin{array}{cc} \nabla_{\boldsymbol{\theta}'}\psi(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) & \nabla_{\boldsymbol{\lambda}'}\psi(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) \end{array}\right] \\ &= \mathbb{E}\left[\begin{array}{cc} \textit{fill} & \textit{this} \\ \textit{in} & \textit{later} \end{array}\right]\end{aligned}$$

13 April 1–15, 2017 – Andrews & Soares (2010)

Our moment equalities from above do not identify α when $\beta = 0$. More generally, the estimator based on them performs poorly when β is relatively small compared to the error variance. Continue to assume that $\alpha_0 = 0$ so the moment conditions simplify.

We now consider an inference procedure following Andrews & Soares. The basic idea is to "isolate" the problematic parameters, in our case α and β, and carry out joint inference for these using the Anderson-Rubin test statistic. This is constructed by substituting a null hypothesis $H_0\colon \theta = \theta_0$ into the sample analogue of the GMM moment conditions and relying on the fact that this sample analogue remains "well-behaved" even in situations where inference for the GMM *parameter estimator* breaks down. Examples include parameters on the boundary, and parameters that may not be identified, e.g. α_1 if $\beta = 0$

13.1 Simple Example: $\alpha_0 = 0$, $c = 0$, and $\sigma_{\varepsilon\varepsilon} = 1$

In our problem c and $\sigma_{\varepsilon\varepsilon}$ are essentially nuisance parameters. Fortunately, they are *always identified* from our moment conditions regardless of the values of β and α, as we will discuss further below. For the moment, we will suppose that c is known to equal zero and $\varepsilon_{\varepsilon\varepsilon}$ is known to equal one as is the case in our baseline simulation. Later we will estimate them which requires only a small modification of the procedure we know outline. With the simplifications $\alpha_0, c = 0, \sigma_{\varepsilon\varepsilon} = 1$ the equality moment conditions become

$$\mathbb{E}\left[\left\{y - \frac{\beta}{1-\alpha_1}T\right\}z\right] = 0$$
$$\mathbb{E}\left[\left\{y^2 - 1 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha_1}T\right\}z\right] = 0$$

since we no longer need the g_1 block of moment conditions to identify the "intercepts" c and $\sigma_{\varepsilon\varepsilon}$. For this simplified set of moment conditions, our parameter vector is $\boldsymbol{\theta} = (\alpha_1, \beta)'$ and the residuals are given by

$$u(\boldsymbol{\theta}) = y - \frac{\beta}{1-\alpha_1}T$$
$$v(\boldsymbol{\theta}) = y^2 - 1 - \frac{\beta}{1-\alpha_1}2yT + \frac{\beta^2}{1-\alpha_1}$$

and we can write the equality moment conditions as

$$\mathbb{E}\left[g(\mathbf{x}, \boldsymbol{\theta})\right] = \mathbb{E}\begin{bmatrix} u(\boldsymbol{\theta})z \\ v(\boldsymbol{\theta})z \end{bmatrix} = \mathbf{0}$$

The inequality moment conditions are unchanged from above, namely

$$\mathbb{E}\left[h(\mathbf{x}, \theta)\right] = \mathbb{E}\begin{bmatrix} (1-\alpha_1) - T(1-z)/(1-q) \\ (1-\alpha_1) - Tz/q \end{bmatrix} \geq \mathbf{0}$$

where $q = \mathbb{E}[z]$. As above we will *condition* on z, i.e. hold it fixed in repeated samples, so we will not add an extra moment condition for q. Instead we will simply substitute the sample analogue. Note that this means we should hold q *fixed* when bootstrapping below. We now introduce some notation from Andrews and Soares (2010).

Population Moment Conditions

$$\mathbb{E}\left[m_j(\mathbf{w}_i, \theta_0)\right] \begin{cases} \geq 0 & \text{for } j = 1, \cdots, p \\ = 0 & \text{for } j = p+1, \cdots, k \text{ where } k = p + v \end{cases}$$

where p is the number of inequality moment conditions (in our case $p = 2$), v is the number of equality moment conditions (in our case $v = 2$), θ_0 is the true parameter vector, and $\mathbf{w}_i$ is the vector of observations for individual i (in our case $\mathbf{w}_i = (T_i, z_i, y_i)$).

Sample Moment Functions, etc.

$$\bar{m}_n(\theta) = \begin{bmatrix} \bar{m}_{n,1}(\theta) \\ \vdots \\ \bar{m}_{n,k}(\theta) \end{bmatrix}, \quad \bar{m}_{n,j} = \frac{1}{n}\sum_{i=1}^{n} m_j(\mathbf{w}_i, \theta) \text{ for } j = 1, \cdots, k$$

Now, let $\Sigma(\theta_0)$ denote the asymptotic variance of $\sqrt{n}\ \bar{m}_n(\theta)$. We estimate this quantity using $\widehat{\Sigma}_n(\theta)$. For iid observations, as in our example, the estimator is

$$\widehat{\Sigma}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[m(\mathbf{w}_i, \theta) - \bar{m}_n(\theta)\right]\left[m(\mathbf{w}_i, \theta) - \bar{m}_n(\theta)\right]', \quad m(\mathbf{w}_i, \theta) = \begin{bmatrix} m_1(\mathbf{w}_i, \theta) \\ \vdots \\ m_k(\mathbf{w}_i, \theta) \end{bmatrix}$$

Test Statistic The test statistic takes the form $T_n(\theta) = S\left(\sqrt{n}\,\bar{m}_n(\theta), \widehat{\Sigma}(\theta)\right)$ for some real-valued function S. The example we will use is S_1, defined by

$$S_1(m, \Sigma) = \sum_{j=1}^{p}[m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v}(m_j/\sigma_j)^2$$

where $m = (m_1, \cdots, m_k)'$,

$$[x]_- = \begin{cases} x, & \text{if } x < 0 \\ 0, & \text{if } x \geq 0 \end{cases}$$

and σ_j^2 is the jth diagonal element of Σ. Notice that S_1 only gives weight to inequality moment conditions that are *violated*.

Basic Idea of the Test Let's return to our specific example for a moment. The idea is essentially to plug a null hypothesis $\theta^* = (\alpha_1^*, \beta^*)'$ into the sample analogue:

$$\sqrt{n}\,\bar{m}_n(\alpha_1^*, \beta^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} (1-\alpha_1^*) - T_i(1-z_i)/(1-q) \\ \\ (1-\alpha_1^*) - T_i z_i/q \\ \\ \left(y_i - \dfrac{\beta^*}{1-\alpha_1^*} T_i\right) z_i \\ \\ \left(y_i^2 - 1 - \dfrac{\beta^*}{1-\alpha_1^*} 2 y_i T_i + \dfrac{\beta^{*2}}{1-\alpha_1^*} T_i\right) z_i \end{bmatrix}$$

and see if the result is "large" after standardizing and squaring the individual elements. We only give weight to an inequality if it is violated. The variance matrix of the sample analogue is calculated *under the null*, i.e. assuming that $\theta = \theta^*$. Note that we also use the *centered variance matrix estimator*.

We reject the null if the test statistic is too large. This gives us *joint inference* for α and β *simultaneously*. To construct a joint confidence region, we need to test pairs (α_1, β). Of course, we restrict α_1 to lie in $[0, 1)$. The resulting confidence region *need not be convex*. In fact it could even be disconnected! However, in our particular example, it might be possible to prove that one gets a connected or even convex region. This is something we should think about since it would reduce the computational burden substantially. To get *marginal* inference, say for β only, one projects the joint confidence set. This is necessarily conservative, but may not be too bad in practice. We'll have to see...

A particularly salient null hypothesis is $\beta = 0$. Imposing this yields

$$\sqrt{n}\,\bar{m}_n(\alpha_1^*, 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} (1-\alpha_1^*) - T_i(1-z_i)/(1-q) \\ (1-\alpha_1^*) - T_i z_i/q \\ y_i z_i \\ (y_i^2 - 1)\, z_i \end{bmatrix}$$

We see that this function depends on α_1^* *only via the moment inequalities*. What is more, the test statistic based on S_1 does *not* depend on α_1^* unless the inequality constraints are violated.

Calculating the Critical Value The test statistic we will use to test $\theta = \theta^*$ is fairly simple to compute: we simply substitute into the GMM sample analogue. The critical value for the test, however, is much more complicated. Following Andrews & Soares (2010), we use the following bootstrap procedure. First we define some additional notation. All of the test statistics considered in Andrews & Soares (2010) satisfy

$$T_n = S\left(\sqrt{n}\,\bar{m}_n(\theta), \widehat{\Sigma}(\theta)\right) = S\left(\widehat{D}^{-1/2}(\theta)\sqrt{n}\,\bar{m}_n(\theta), \widehat{\Omega}_n(\theta)\right)$$

where

$$\widehat{D}_n(\theta) = \text{diag}\left(\widehat{\Sigma}\,(\theta)\right), \quad \widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\,\widehat{\Sigma}(\theta)\,\widehat{D}_n^{-1/2}(\theta)$$

Now, let $\{\mathbf{w}_i^*\}_{i=1}^n$ denote a bootstrap sample and define the associated bootstrap quantities

$$\begin{aligned}
M_n^*(\theta) &= \sqrt{n}\left(\widehat{D}^*\,(\theta)\right)^{-1/2}\left(\bar{m}_n^*\,(\theta) - \bar{m}_n\,(\theta)\right)) \\
\widehat{\Omega}^*(\theta) &= \left(\widehat{D}^*\,(\theta)\right)^{-1/2}\widehat{\Sigma}_n^*(\theta)\left(\widehat{D}^*\,(\theta)\right)^{-1/2} \\
\widehat{D}^*\,(\theta) &= \text{diag}\left(\widehat{\Sigma}_n^*\,(\theta)\right) \\
m_n^*(\theta) &= \frac{1}{n}\sum_{i=1}^n m(\mathbf{w}_i^*, \theta) \\
\widehat{\Sigma}(\theta)^* &= \frac{1}{n}\sum_{i=1}^n \left[m(\mathbf{w}_i^*, \theta) - \bar{m}_n^*(\theta)\right]\left[m(\mathbf{w}_i^*, \theta) - \bar{m}_n^*(\theta)\right]'
\end{aligned}$$

Note that $M_n^*(\theta)$ is centered around the *non-bootstrap* sample analogue $\bar{m}_n(\theta)$: this is *very important!* Now we describe the procedure for calculating the bootstrap critical value:

1. Calculate $\sqrt{n}\,\bar{m}_n\,(\theta_0)$ and $\widehat{\Sigma}(\theta_0)$ under the null hypothesis $H_0\colon \theta = \theta_0$.

2. Determine which inequality moment conditions are "far from binding" as follows:

    - Let $j \in J = \{1, \cdots, p\}$ index the inequality moment conditions.
    - Let $\widehat{\sigma}_{n,j}(\theta_0)^2$ denote the (j, j) element of $\widehat{\Sigma}(\theta_0)$
    - For each $j \in J$ calculate the "t-statistic" $t_{n,j} = \sqrt{n}\,\bar{m}_j(\theta_0)/\widehat{\sigma}_{n,j}(\theta_0)$
    - Let $\mathcal{FB}$ denote the subset of J for which $t_{n,j} > \sqrt{\log n}$. These are the inequality moment conditions that are "far from binding" under $H_0\colon \theta = \theta_0$.

3. Calculate the test statistic $T_n = S_1\left(\sqrt{n}\,\bar{m}_n\,(\theta_0)\,, \widehat{\Sigma}\,(\theta_0)\right)$.

4. Calculate the bootstrap critical value for the test as follows:

    - Draw R bootstrap samples – each with sample size n.

- For each bootstrap sample, r, calculate $M^{**}_{n,r}(\theta_0)$ and $\widehat{\Omega}^{**}_{n,r}(\theta_0)$ – the bootstrap versions of $M_n(\theta_0)$ and $\widehat{\Omega}(\theta_0)$, defined above but with a slight change: *drop* any moment inequality $j \in \mathcal{FB}$. That is, drop any inequality that we determined was far from binding *on the basis of the real data* (i.e. *not* this bootstrap sample!)
- For each bootstrap sample r calculate $T^{**}_{n,r} = S_1\left(M^{**}_{n,r}(\theta_0), \widehat{\Omega}^{**}_{n,r}(\theta_0)\right)$.
- Set $\widehat{c}_n(\theta_0, 1-\delta)$ equal to the $1-\delta$ sample quantile of the $\left\{T^{**}_{n,r}\right\}_{r=1}^{R}$

5. Reject $H_0\colon \theta = \theta_0$ if $T_n > \widehat{c}_n(\theta_0, 1-\delta)$
6. To construct a $(1-\delta)\times 100\%$ confidence set, invert the test of $H_0\colon \theta = \theta_0$ for $\theta_0 \in \Theta$.

14 April 15–18, 2017

Second Moment Inequalities

Step 1: Second Moment Bounds Our model has $\varepsilon = y - c - bT^*$ where $\mathbb{E}[\varepsilon] = 0$. Conditional on $T^* = 0$, $\varepsilon = y - c$ and conditional on $T^* = 1$, $\varepsilon = y - c - \beta$. Hence,

$$\begin{aligned}
\mathbb{E}\left[\varepsilon^2 | T^* = 0, z = k\right] &= \mathbb{E}\left[y^2 - 2cy | T^* = 0, z = k\right] + c^2 \\
\mathbb{E}\left[\varepsilon^2 | T^* = 1, z = k\right] &= \mathbb{E}\left[y^2 - 2(\beta + c)y | T^* = 1, z = k\right] + (\beta + c)^2
\end{aligned}$$

The bounds will impose that $\text{Var}(\varepsilon | T^*, z) > 0$ which is equivalent to $\mathbb{E}\left[\varepsilon^2 | T^*, z\right] > 0$ since ε is mean zero:

$$\begin{aligned}
\mathbb{E}\left[y^2 | T^* = 0, z = k\right] &> 2c\mathbb{E}\left[y | T^* = 0, z = k\right] - c^2 \\
\mathbb{E}\left[y^2 | T^* = 1, z = k\right] &> 2(\beta + c)\mathbb{E}\left[y | T^* = 1, z = k\right] - (\beta + c)^2
\end{aligned}$$

These are not in fact the bounds we want! I messed up since ε is *not* mean zero *conditional* on T^* and z. These bounds are still correct, but they're not the best we can do and may not in fact be better than the simple first-moment bounds...

Step 2: Relate $\mathbb{E}[y^r | T^*, z]$ to $\mathbb{E}[y^r | T, z]$ By iterated expectations and the assumption that $\mathbb{E}\left[y^r | T, T^*, z\right] = \mathbb{E}\left[y^r | T^*, z\right]$,

$$\begin{aligned}
\mathbb{E}\left[y^r | T = 0, z_k\right] &= \mathbb{E}\left[y^r | T^* = 0, z_k\right]\mathbb{P}(T^* = 0 | T = 0, z_k) + \mathbb{E}\left[y^r | T^* = 1, z_k\right]\mathbb{P}(T^* = 1 | T = 0, z_k) \\
\mathbb{E}\left[y^r | T = 1, z_k\right] &= \mathbb{E}\left[y^r | T^* = 0, z_k\right]\mathbb{P}(T^* = 0 | T = 1, z_k) + \mathbb{E}\left[y^r | T^* = 1, z_k\right]\mathbb{P}(T^* = 1 | T = 1, z_k)
\end{aligned}$$

The preceding is a linear system of the form

$$
\begin{aligned}
a &= \pi x + (1-\pi) y \\
b &= (1-\delta) x + \delta y
\end{aligned}
$$

and hence its solution is

$$
\begin{aligned}
x &= \left[\frac{\delta}{\pi\delta - (1-\pi)(1-\delta)}\right] a + \left[\frac{-(1-\pi)}{\pi\delta - (1-\pi)(1-\delta)}\right] b \\
y &= \left[\frac{-(1-\delta)}{\pi\delta - (1-\pi)(1-\delta)}\right] a + \left[\frac{\pi}{\pi\delta - (1-\pi)(1-\delta)}\right] b
\end{aligned}
$$

by Bayes' rule, as we show in the appendix to sick-instruments,

$$
\begin{aligned}
\pi &= \mathbb{P}(T^* = 0 | T = 0, z_k) = (1-\alpha_0)(1-p_k^*)/(1-p_k) \\
1-\pi &= \mathbb{P}(T^* = 1 | T = 0, z_k) = \alpha_1 p_k^*/(1-p_k) \\
\delta &= \mathbb{P}(T^* = 1 | T = 1, z_k) = (1-\alpha_1) p_k^*/p_k \\
1-\delta &= \mathbb{P}(T^* = 0 | T = 1, z_k) = \alpha_0 (1-p_k^*)/p_k
\end{aligned}
$$

Some algebra shows that

$$
\pi\delta - (1-\pi)(1-\delta) = \frac{(p_k - \alpha_0)(1 - p_k - \alpha_1)}{1 - \alpha_0 - \alpha_1}
$$

Hence, after simplifying and rearranging, it follows that

$$
\begin{aligned}
p_k(1-p_k)(1-p_k-\alpha_1)\mathbb{E}\left[y^r | T^* = 0, z_k\right] &= (1-\alpha_1)(1-p_k)\mathbb{E}[y^r | T = 0, z_k] - \alpha_1 p_k \mathbb{E}[y^r | T = 1, z_k] \\
p_k(1-p_k)(p_k - \alpha_0)\mathbb{E}\left[y^r | T^* = 1, z_k\right] &= (1-\alpha_0) p_k \mathbb{E}[y^r | T = 1, z_k] - \alpha_0(1-p_k)\mathbb{E}[y^r | T = 0, z_k]
\end{aligned}
$$

Step 3: Convert Conditional to Unconditional Moments By iterated expectations and the assumption that $\mathbb{E}\left[y^r | T, T^*, z\right] = \mathbb{E}\left[y^r | T^*, z\right]$

$$
\begin{aligned}
p_k &= \mathbb{E}[T | z = k] = \mathbb{E}\left[T\mathbf{1}(z = k)\right] / \mathbb{P}(z = k) \\
\mathbb{E}\left[y^r | T = 0, z = k\right] &= \mathbb{E}\left[y^r(1-T)\mathbf{1}(z = k)\right] / \left[(1-p_k)\mathbb{P}(z = k)\right] \\
\mathbb{E}\left[y^r | T = 1, z = k\right] &= \mathbb{E}\left[y^r T\mathbf{1}(z = k)\right] / \left[p_k \mathbb{P}(z = k)\right]
\end{aligned}
$$

Step 4: Substitute Steps 2–3 into Step 1 After some algebra, we find that

$$\mathbb{E}\left[\mathbf{1}(z=k)\left\{T-(1-\alpha_1)\right\}\left(y^2-2cy\right)\right] < c^2\mathbb{P}(z=k)p_k(1-p_k)(1-p_k-\alpha_1)$$
$$\mathbb{E}\left[\mathbf{1}(z=k)(\alpha_0-T)\left\{y^2-2(\beta+c)y\right\}\right] < (\beta+c)^2\mathbb{P}(z=k)p_k(1-p_k)(p_k-\alpha_0)$$

Step 5: Complete the Square

$$\mathbb{E}\left[\mathbf{1}(z=k)\left\{1-\alpha_1-T\right\}(y-c)^2\right] > c^2\mathbb{P}(z=k)(1-p_k-\alpha_1)\left[1-p_k(1-p_k)\right]$$
$$\mathbb{E}\left[\mathbf{1}(z=k)(T-\alpha_0)\left\{y-(\beta+c)\right\}^2\right] > (\beta+c)^2\mathbb{P}(z=k)(p_k-\alpha_0)\left[1-p_k(1-p_k)\right]$$

15 April 21–23

The Full Set of Moment Inequalities

Above we considered the special case in which $\alpha_0 = 0$ so that first and second moments were sufficient to identify the model. We now derive the GMM-style moment conditions for the general case in which α_0 may not equal zero and we need to use third moments for identification.

Note that we have checked all of these equalities numerically in our simulation study and they are indeed correct!

Notation and Identification Define the following re-parameterization

$$\begin{aligned}
\theta_1 &= \beta/(1-\alpha_0-\alpha_1)\\
\theta_2 &= \theta_1^2\left[1+(\alpha_0-\alpha_1)\right]\\
\theta_3 &= \theta_1^3\left[(1-\alpha_0-\alpha_1)^2+6\alpha_0\left(1-\alpha_1\right)\right]
\end{aligned}$$

Using this notation, the covariance equations from above become

$$\begin{aligned}
\text{Cov}(y,z)-\theta_1\text{Cov}(T,z) &= 0\\
\text{Cov}(y^2,z)-\theta_1 2\text{Cov}(yT,z)+\theta_2\text{Cov}(T,z) &= 0\\
\text{Cov}(y^3,z)-\theta_1 3\text{Cov}(y^2T,z)+\theta_2 3\text{Cov}(yT,z)-\theta_3\text{Cov}(T,z) &= 0
\end{aligned}$$

Note that it is trivial to prove that θ_1, θ_2 and θ_3 are identified. To show that β, α_0 and α_1 are identified, write

$$\theta_2/\theta_1^2 = 1 + (\alpha_0 - \alpha_1)$$
$$\theta_3/\theta_1^3 = (1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1)$$

which we can always do provided that $\theta_1 \neq 0$ which is equivalent to $\beta \neq 0$. Re-arranging the first equation allows us to solve for α_0 as a function of α_1. Substituting this into the second gives us a quadratic in α_1 only. This should be identical to the quadratic we obtained in our original identification proof.

First Moment Equalities Expanding,

$$\begin{aligned} \text{Cov}(y, z) - \theta_1 \text{Cov}(T, z) &= 0 \\ \mathbb{E}[yz - \theta_1 Tz] - \mathbb{E}[z]\mathbb{E}[y - \theta_1 T] &= 0 \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[y - \theta_1 T] &= c + \beta \mathbb{E}\left[T^* - T/(1 - \alpha_0 - \alpha_1)\right] \\ &= c + \beta \left\{ \frac{\mathbb{E}[T] - \alpha_0}{1 - \alpha_0 - \alpha_1} - \frac{\mathbb{E}[T]}{1 - \alpha_0 - \alpha_1} \right\} \\ &= c - \alpha_0 \theta_1 \end{aligned}$$

Thus,

$$\mathbb{E}\left[\{y - (c - \alpha_0 \theta_1) - \theta_1 T\} z\right] = 0$$

Now define

$$\kappa_1 = c - \alpha_0 \theta_1$$
$$u_1 = y - \kappa_1 - \theta_1 T$$

Using this notation, we obtain the unconditional moment equalities

$$\mathbb{E}\left[\begin{array}{c} u_1(\kappa_1, \theta_1) \\ u_1(\kappa_1, \theta_1) z \end{array}\right] = \mathbf{0}$$

Second Moment Equalities

$$\begin{aligned}\mathrm{Cov}(y^2,z)-\theta_1 2\mathrm{Cov}(yT,z)+\theta_2\mathrm{Cov}(T,z)&=0\\ \mathbb{E}\left[y^2z-\theta_1 2yTz+\theta_2 Tz\right]-\mathbb{E}\left[z\right]\mathbb{E}\left[y^2-\theta_1 2yT+\theta_2 T\right]&=0\end{aligned}$$

Now, using some lemmas from the notes in our sick-instruments paper,

$$\begin{aligned}\mathbb{E}[T\varepsilon]&=\mathrm{Cov}(T,\varepsilon)\\ &=\mathrm{Cov}(T^*,\varepsilon)+\mathrm{Cov}(w,\varepsilon)\\ &=\mathrm{Cov}(T^*,\varepsilon)-\mathrm{Cov}(T^*,\varepsilon)(\alpha_0+\alpha_1)\\ &=\mathbb{E}(T^*\varepsilon)(1-\alpha_0-\alpha_1)\end{aligned}$$

hence,

$$\begin{aligned}\mathbb{E}[yT]&=c\mathbb{E}[T]+\beta\mathbb{E}[TT^*]+\mathbb{E}\left[T,\varepsilon\right]\\ &=cp+\beta\mathbb{P}(T=1,T^*=1)+\mathbb{E}[T^*\varepsilon](1-\alpha_0-\alpha_1)\\ &=cp+\beta(1-\alpha_1)p^*+\mathbb{E}[T^*\varepsilon](1-\alpha_0-\alpha_1)\end{aligned}$$

Combining this with

$$\mathbb{E}[y^2]=c^2+\beta^2p^*+\sigma_{\varepsilon\varepsilon}+2c\beta p^*+2\beta\mathbb{E}[T^*\varepsilon]$$

we find that

$$\mathbb{E}\left[y^2-\theta_1 2yT+\theta_2 T\right]=\cdots=c^2+\sigma_{\varepsilon\varepsilon}+\alpha_0(\theta_2-2c\theta_1)$$

For the steps, see our whiteboard notes from 2017-04-21 17.06.32

Now, define

$$\begin{aligned}\kappa_2&=c^2+\sigma_{\varepsilon\varepsilon}+\alpha_0(\theta_2-2c\theta_1)\\ u_2&=y^2-\kappa_2-\theta_1 2yT+\theta_2 T\end{aligned}$$

Using this notation, we obtain the unconditional moment equalities

$$\mathbb{E}\begin{bmatrix}u_2(\kappa_2,\theta_1,\theta_2)\\ u_2(\kappa_2,\theta_1,\theta_2)z\end{bmatrix}=\mathbf{0}$$

Third Moment Equalities

$$\begin{aligned}\text{Cov}(y^3, z) - \theta_1 3\text{Cov}(y^2T, z) + \theta_2 3\text{Cov}(yT, z) - \theta_3 \text{Cov}(T, z) = 0\\ \mathbb{E}\left[y^3 z - \theta_1 3y^2Tz + \theta_2 3yTz - \theta_3 Tz\right] - \mathbb{E}[z]\mathbb{E}\left[y^3 - \theta_1 3y^2T + \theta_2 3yT - \theta_3 T\right] = 0\end{aligned}$$

$$\mathbb{E}[y^3] = \cdots = c^3 + \beta p^*\left(3c^2 + 3c\beta + \beta^2\right) + 3\beta\mathbb{E}[\varepsilon T^*](2c+\beta) + 3c\sigma_{\varepsilon\varepsilon} + 3\beta\mathbb{E}[\varepsilon^2 T^*] + \mathbb{E}[\varepsilon^3]$$

For derivations of the following, see our whiteboard notes from April 21st and 22nd, 2017

$$\mathbb{E}\left[y^2T\right] = \cdots = c^2p + \beta(1-\alpha_1)p^* + \mathbb{E}[T\varepsilon^2] + 2c\beta(1-\alpha_1)p^* + 2c\mathbb{E}[T\varepsilon] + 2\beta\mathbb{E}[TT^*\varepsilon]$$

and

$$\begin{aligned}\mathbb{E}[T\varepsilon^2] &= \cdots = \alpha_0\sigma_{\varepsilon\varepsilon} + (1-\alpha_0-\alpha_1)\mathbb{E}[\varepsilon^2T^*]\\ \mathbb{E}[TT^*\varepsilon] &= \cdots = (1-\alpha_1)\mathbb{E}[T^*\varepsilon]\end{aligned}$$

We then need to calculate $\mathbb{E}[y^3 - \theta_1 3y^2T + \theta_2 3yT - \theta_3 T]$. Using the preceding expressions, we can show after some algebra that the $\mathbb{E}[T^*\varepsilon]$ and $\mathbb{E}[T^*\varepsilon^2]$ terms drop out, hence

$$\begin{aligned}\mathbb{E}[y^3 - \theta_1 3y^2T + \theta_2 3yT - \theta_3 T] = &\left\{c^3 + \beta p^*\left(3c^2 + 3c\beta + \beta^2\right)\right\} + 3c\sigma_{\varepsilon\varepsilon} + \mathbb{E}[\varepsilon^3] - \theta_1 3\alpha_0\sigma_{\varepsilon\varepsilon} - \theta_3 p\\ &- \theta_1 3\left\{c^2p + \beta^2(1-\alpha_1)p^* + 2c\beta(1-\alpha_1)p^*\right\} + \theta_2 3\left\{cp + \beta(1-\alpha_1)p^*\right\}\end{aligned}$$

After *even more algebra* we can show that this expression depends neither on p nor on p^*:

$$\begin{aligned}\mathbb{E}[y^3 - \theta_1 3y^2T + \theta_2 3yT - \theta_3 T] &= c^3 + 3\left(c - \theta_1\alpha_0\right)\sigma_{\varepsilon\varepsilon} + \mathbb{E}[\varepsilon^3] - \alpha_0\theta_3 - 3c\alpha_0\left[\theta_1\left(c+\beta\right) - 2\theta_2\frac{1-\alpha_1}{1+\alpha_0-\alpha_1}\right]\\ &= c^3 + 3\left(c - \theta_1\alpha_0\right)\sigma_{\varepsilon\varepsilon} + \mathbb{E}[\varepsilon^3] - \alpha_0\theta_3 - 3c\alpha_0\left[\theta_1\left(c+\beta\right) - 2\theta_1^2(1-\alpha_1)\right]\end{aligned}$$

Now let

$$\begin{aligned}\kappa_3 &= c^3 + 3\left(c - \theta_1\alpha_0\right)\sigma_{\varepsilon\varepsilon} + \mathbb{E}[\varepsilon^3] - \alpha_0\theta_3 - 3c\alpha_0\left[\theta_1\left(c+\beta\right) - 2\theta_1^2(1-\alpha_1)\right]\\ u_3 &= y^3 - \kappa_3 - \theta_1 3y^2T + \theta_2 3yT - \theta_3 T\end{aligned}$$

we obtain the following moment equalities

$$\mathbb{E}\left[\begin{array}{c} u_3(\kappa_3, \theta_1, \theta_2, \theta_3) \\ u_3(\kappa_3, \theta_1, \theta_2, \theta_3) z \end{array}\right] = \mathbf{0}$$

Putting Everything Together Define

$$\begin{aligned} \boldsymbol{\kappa} &= (\kappa_1, \kappa_2, \kappa_3)' \\ \boldsymbol{\theta} &= (\theta_1, \theta_2, \theta_3)' \\ \mathbf{u}(\boldsymbol{\kappa}, \boldsymbol{\theta}) &= \left[\begin{array}{ccc} u_1(\kappa_1, \theta_1) & u_2(\kappa_2, \theta_2) & u_3(\kappa_3, \theta_3) \end{array}\right]' \end{aligned}$$

where

$$\begin{aligned} \kappa_1 &= c - \alpha_0\theta_1 \\ \kappa_2 &= c^2 + \sigma_{\varepsilon\varepsilon} + \alpha_0(\theta_2 - 2c\theta_1) \\ \kappa_3 &= c^3 + 3\left(c - \theta_1\alpha_0\right)\sigma_{\varepsilon\varepsilon} + \mathbb{E}[\varepsilon^3] - \alpha_0\theta_3 - 3c\alpha_0\left[\theta_1\left(c + \beta\right) - 2\theta_1^2(1 - \alpha_1)\right] \end{aligned}$$

and

$$\begin{aligned} u_1(\kappa_1, \theta_1) &= y - \kappa_1 - \theta_1 T \\ u_2(\kappa_2, \theta_1, \theta_2) &= y^2 - \kappa_2 - \theta_1 2yT + \theta_2 T \\ u_3(\kappa_3, \theta_1, \theta_2, \theta_3) &= y^3 - \kappa_3 - \theta_1 3y^2 T + \theta_2 3yT - \theta_3 T \end{aligned}$$

Then the full system of moment equalities is given by

$$\mathbb{E}\left[\begin{array}{c} \mathbf{u}\left(\boldsymbol{\kappa}, \boldsymbol{\theta}\right) \\ \mathbf{u}\left(\boldsymbol{\kappa}, \boldsymbol{\theta}\right) z \end{array}\right] = \mathbf{0}$$

where

$$\begin{aligned} \theta_1 &= \beta/(1 - \alpha_0 - \alpha_1) \\ \theta_2 &= \theta_1^2\left[1 + (\alpha_0 - \alpha_1)\right] \\ \theta_3 &= \theta_1^3\left[(1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0\left(1 - \alpha_1\right)\right] \end{aligned}$$

What about exogenous covariates? At one extreme, we could simply carry out the above *conditional* on $\mathbf{x}$. This would entail treating $\boldsymbol{\kappa}$ and $\boldsymbol{\theta}$ as functions of $\mathbf{x}$ such that α_0, α_1 and β would be allowed to depend on $\mathbf{x}$. To implement this, one would need to estimate a number of conditional mean functions: $\mathbb{E}[y|\mathbf{x}]$, $\mathbb{E}[T|\mathbf{x}]$, $\mathbb{E}[yT|\mathbf{x}]$, $\mathbb{E}[y^3|\mathbf{x}]$, and $\mathbb{E}[y^2T|\mathbf{x}]$. It would appear that this leads to a standard two-step estimation problem although clearly you'd need a lot of data to have any chance! Another idea would be to impose some restrictions on the way in which $\mathbf{x}$ affects α_0 etc, leading to a semi-parametric estimator. At the other extreme, we could impose the assumption that $\mathbf{x}$ affects y linearly and α_0, α_1 do not vary with $\mathbf{x}$:

$$y = c + \beta T^* + \mathbf{x}'\boldsymbol{\gamma} + \varepsilon$$

Under this assumption the result of Frazis & Loewenstein (2003) holds – the IV estimator would recover $\kappa_1, \theta_1, \boldsymbol{\gamma}$. Since $(y - \mathbf{x}'\boldsymbol{\gamma}) = c + \beta T^* + \varepsilon$, all of the preceding moment conditions should still hold only with $(y - \mathbf{x}'\boldsymbol{\gamma})$ in place of y. To identify $\boldsymbol{\gamma}$ we add an extra moment condition in which $\mathbf{x}$ multiplies the u_1 error term, re-defined to subtract $\mathbf{x}'\boldsymbol{\gamma}$.

16 May 5th, 2017 – Todo for Revised Paper

1. New notation to accomodate covariates etc. as in Mahajan and Lewbel

2. Move Mahajan stuff into appendix; possibly convert back to his notation later – I have handwritten notes on this.

3. Identification results: identification from higher moments, lack of identification from conditional means alone

4. Explain briefly how to do fully non-parametric estimation, specialize to some simple cases, e.g. linear model

5. Explain about inference versus estimation.

6. Possibly show that Mahajan, Lewbel, etc. also suffer from a weak identification problem. This might be helpful for selling the paper. Would need to think about what inequalities to use in their case. Presumably we could still use the "weak" bounds, but there might be others we could exploit.

7. Testing $\beta = 0$ should be very easy: don't need Andrews & Soares at all. I think we can either just use the Stock & Wright GMM-AR or even a plain old-fashioned GMM inference for testing a linear restriction.

8. Canay et al. projection inference for β? Need to figure out if this actually gives a speed-up in our case. If so, could be very useful for simulations. Would be nice to

show that we have power to reject the probability limit of the IV and reduced form. Is it easier to handle the strongly-identified parameters using their method?

9. Could also talk briefly about the continuum of moment conditions idea and how one obtains identification from quantiles. There's also a question of using more inequalities by imposing the independence assumption for the measurement error. Could also proceed without our higher moment assumptions and just get inference for the identified set.

10. Is there any way to handle treatment effect heterogeneity? Should be ok if it's modeled heterogeneity but maybe there are some other special cases we can handle? What about that quantile treatment effect idea?

11. Simulations: want to show that we can actually learn something useful. Things we want to show are that we have more power than just the RF for testing $\beta = 0$, that we have power to reject the RF and IV plims in certain cases.

12. Need to show at least in simulations that our confidence regions aren't insane. Would be good if we could formally argue that they must be convex and connected, for example. Maybe the key is to show that our inequalities are always well-behaved and that adding the equalities cann't make the problem suddenly become badly behaved.

13. Andrews & Soares with estimates of the strongly identified parameters to get a joint region for α_0, α_1, and β. But this is slow, so we probably can't do simulations although it would be interesting in the empirical examples.

14. Try to get the projection inference for β working

15. Empirical example or examples: Oreopolous? Maybe the Heckman JPTA dataset? Petra suggested emailing someone about this but I forget who it was...

17 Lewbel (2007)

Notation Observe Y, Z and T where T is a proxy for T^* and

$$
\begin{aligned}
h^*(X, T^*) &= E(Y|X, T^*) \\
T^* &= \text{unobserved binary regressor} \\
X &= \text{vector of covariates} \\
Y &= \text{outcome}
\end{aligned}
$$

Since T^* is binary, without loss of generality,

$$\begin{aligned} h^*(X, T^*) &= h_0^*(X) + \tau^*(x)T^* \\ h_0^*(X) &= h^*(X, 0) \\ \tau^*(x) &= h^*(x, 1) - h^*(x, 0) \end{aligned}$$

Goal is to estimate $\tau^*(x)$. Below we will partition X according to $X = (V, Z)$ where V is an "instrument-like variable" and Z is the set of remaining covariates.

Assumption A1 There exists $E(Y|X, T^*, T) = E(Y|X, T^*)$. Equivalently, Y is mean-independent of $T - T^*$, conditional on X, T^* – mis-classification does not affect true expected outcome. This rules out placebo effects, etc.

More Notation

$$\begin{aligned} r^*(x) &= E(T^*|X = x) = P(T^* = 1|X = x) \\ b_0(x) &= P(T = 1|T^* = 0, X = x) \\ b_1(x) &= P(T = 0|T^* = 1, X = x) \\ r(x) &= E(T|X = x) \\ \tau(x) &= h(x, 1) - h(x, 0) \\ h(x, t) &= E(Y|X = x, T = t) \end{aligned}$$

Assumption A2 There exist $b_0(x) + b_1(x) < 1$ and $0 < r^*(x) < 1$ for all x in the support of X.

Theorem 1 Under Assumption A1 there exists a function $m(x)$ such that $|m(x)| \leq 1$ and $\tau(x) = \tau^*(x)m(x)$. If we add Assumption A2 then $m(x) > 0$ as well.

- Analogous to attenuation bias under classical measurement error: A1 implies that $\tau(x)$ is a lower bound for the magnitude of $\tau^*(x)$. Adding A2, sign of $\tau(x)$ agrees with that of $\tau^*(x)$.

- Assumptions A1 and A2 imply that $\tau^*(x) = 0$ iff $\tau(x) = 0$. Thus, if we only want to test $\tau^*(x) = 0$ we can simply ignore mis-classification.

- As shown below:

$$m(x) = P(T^* = 1|T = 1, X = x) - P(T^* = 1|T = 0, X = x) \implies |m(x)| < 1$$
$$m(x) = M[b_0(x), b_1(x), r(x)] = \frac{1}{1 - b_1(x) - b_0(x)} \left\{ 1 - \frac{[1 - b_1(x)]\, b_0(x)}{r(x)} - \frac{[1 - b_0(x)]\, b_1(x)}{1 - r(x)} \right\}$$
$$m(x) = \frac{[1 - r^*(x)] r^*(x)}{[1 - r(x)]\, r(x)} [1 - b_0(x) - b_1(x)] \implies m(x) > 0 \text{ if } 1 - b_0(x) + b_1(x) > 0$$

Proof of Theorem 1. Let $p_t(X) = P(T^* = 1|X, T = t)$. By A1 we have

$$E(Y|X, T^*, T) = E(Y|X, T^*) = h_0^*(X) + \tau^*(X)T^*$$

combining this with iterated expectations,

$$\begin{aligned} E(Y|X, T = t) &= E_{T^*|X,T=t}\left[E(Y|X, T = t, T^*)\right] = E_{T^*|X,T=t}\left[h_0^*(X) + \tau^*(X)T^*\right] \\ &= p_t(X)\left[h_0^*(X) + \tau^*(X)\right] + \left[1 - p_t(X)\right] h_0^*(X) \\ &= h_0^*(X) + p_t(X)\tau^*(X) \end{aligned}$$

Taking the difference of the preceding expression evaluated at $T = 1$ and $T = 0$

$$E(Y|X, T = 1) - E(Y|X, T = 0) = [p_1(X) - p_0(X)]\, \tau^*(X)$$

whereas we defined

$$\tau(X) = h(X, 1) - h(X, 0) = E(Y|X, T = 1) - E(Y|X, T = 0)$$

Combining these two equations, we see that

$$\tau(X) = [p_1(X) - p_0(X)]\, \tau^*(X)$$

so that the function $m(x)$ defined in Theorem 1 equals $p_1(x) - p_0(x)$. Since m is a difference of probabilities, it follows immediately that $-1 \leq m(x) \leq 1$. Now, by Bayes' Rule,

$$p_0(x) = P(T^* = 1|X = x, T = 0) = \frac{P(T = 0|X = x, T^* = 1)P(T^* = 1|X = x)}{P(T = 0|X = x)} = \frac{b_1(x)r^*(x)}{1 - r(x)}$$
$$p_1(x) = P(T^* = 1|X = x, T = 1) = \frac{P(T = 1|X = x, T^* = 1)P(T^* = 1|X = x)}{P(T = 1|X = x)} = \frac{[1 - b_1(x)]r^*(x)}{r(x)}$$

and by iterated expectations,

$$\begin{aligned}
r(x) &= E(T|X=x) = E(T|X=x, T^*=1)P(T^*=1|X=x) + E(T|X=x, T^*=0)P(T^*=0|X=x)\\
&= [1-b_1(x)]r^*(x) + b_0(x)[1-r^*(x)] = b_0(x) + r^*(x)[1-b_0(x)-b_1(x)]
\end{aligned}$$

Using this expression, if $b_0(x)+b_1(x)=1$ then $r(x)=b_0$. If instead $b_0(x)+b_1(x)\neq 1$, then we can divide through by $1-b_0(x)-b_1(x)$ to solve for $r^*(x)$ and $1-r^*(x)$ as follows:

$$\begin{aligned}
r^*(x) &= \frac{r(x)-b_0(x)}{1-b_0(x)-b_1(x)}\\
1-r^*(x) &= \frac{[1-b_0(x)-b_1(x)]-[r(x)-b_0(x)]}{1-b_0(x)-b_1(x)} = \frac{1-b_1(x)-r(x)}{1-b_0(x)-b_1(x)}
\end{aligned}$$

Using the expressions we have just derived for p_0, p_1, r^* and $1-r^*$ and suppressing the dependence on x for simplicity, it follows that

$$\begin{aligned}
m = p_1 - p_0 &= \frac{(1-b_1)r^*}{r} - \frac{b_1 r^*}{1-r} = \frac{(1-r)(1-b_1)r^* - rb_1r^*}{r(1-r)}\\
&= \frac{r^*(1-r-b_1+rb_1-rb_1)}{r(1-r)} = \frac{r^*(1-r-b_1)}{r(1-r)}
\end{aligned}$$

Rearranging,

$$(1-r)rm = r^*(1-r-b_1)$$

and combining this with $1-r^* = (1-b_1-r)/(1-b_0-b_1)$, we have

$$(1-r)rm = (1-r^*)r^*(1-b_0-b_1)$$

Both r and r^* are strictly between zero and one. Thus, if $b_0+b_1=1$ we have $m=0$. If instead $b_0+b_1<1$ we have $m>0$ and if $b_0+b_1>1$ then $m<0$. Finally,

$$m = \frac{r^*(1-r-b_1)}{r(1-r)} = \frac{r-b_0}{1-b_0-b_1}\left[\frac{1-r-b_1}{r(1-r)}\right]$$

and since

$$\begin{aligned}(r-b_0)(1-r-b_1) &= r(1-r) - b_0(1-r) - b_1 r + b_1 b_0 \\ &= r(1-r) - b_0(1-r) - b_1 r + b_1 b_0 + (b_1 b_0 r - b_1 b_0 r) \\ &= r(1-r) - b_0(1-r) - b_1 r + b_1 b_0(1-r) + b_1 b_0 r \\ &= r(1-r) - b_0(1-r) - b_1 r(1-b_0) + b_1 b_0(1-r) \\ &= r(1-r) + b_0(1-r)(b_1-1) - b_1 r(1-b_0) \\ &= r(1-r) - b_0(1-r)(1-b_1) - b_1 r(1-b_0)\end{aligned}$$

we find that

$$\begin{aligned}m = M(b_0, b_1, r) &= \frac{1}{1-b_0-b_1}\left[\frac{r(1-r) - (1-b_1)b_0(1-r) - (1-b_0)b_1 r}{r(1-r)}\right] \\ &= \frac{1}{1-b_0-b_1}\left[1 - \frac{(1-b_1)b_0}{r} - \frac{(1-b_0)b_1}{1-r}\right].\end{aligned}$$

□

Assumption A3 Assume that $r(x)$ and $\tau(x)$ are identified. Note that this only requires that we can consistently estimate conditional expectations of observable quantities.

Assumption A4 Suppose that we have partitioned X into two subvectors: $X = (V, Z)$. We assume that for each z in the support of Z there exists a subset Ω_z of the support of V such that:

(i) for all $v, v' \in \Omega_z$, $b_0(v,z) = b_0(v',z)$, $b_1(v,z) = b_1(v',z)$, and $\tau^*(v,z) = \tau^*(v',z)$

(ii) for all $v, v' \in \Omega_z$ such that $v \neq v'$, $r^*(v,z) \neq r^*(v',z)$.

The basic idea here is that V affects the probability of being treated r^* but not the treatment effect τ^* after we have conditioned on Z. As sufficient condition for the restriction on τ^* to hold is $E(Y|Z=z, V=v, T^*=t) = s_1(z,t) + s_2(z,v)$ for some functions s_1 and s_2.

Some More Notation Let $b_0(z), b_1(z), \tau^*(z)$ denote $b_0(v,z), b_1(v,z), \tau^*(v_z)$ for $v \in \Omega_z$ since, under A4, these do not vary with v.

Assumption A5 Each set Ω_z from A4 contains three elements v_k, $k = 0, 1, 2$ such that

$$\left[\frac{\tau(v_0,z)}{r(v_1,z)} - \frac{\tau(v_1,z)}{r(v_0,z)}\right]\left[\frac{\tau(v_0,z)}{1-r(v_2,z)} - \frac{\tau(v_2,z)}{1-r(v_0,z)}\right] \neq \left[\frac{\tau(v_0,z)}{r(v_2,z)} - \frac{\tau(v_2,z)}{r(v_0,z)}\right]\left[\frac{\tau(v_0,z)}{1-r(v_1,z)} - \frac{\tau(v_1,z)}{1-r(v_0,z)}\right]$$

- The key requirement is that V takes on at least three values.
- Assumption A5 depends only on observables, so we can test it.
- An equivalent way to state A5 is: $\tau^*(z) \neq 0$, $b_0(z) + b_1(z) \neq 1$ and an inequality involving r and r^* which is explained below.
- The triplets (v_0, v_1, v_2) are allowed to depend on z.

Theorem 2 Under A1–A5, the mis-classification probabilities $b_0(x), b_1(x)$ are identified as is the probability of treatment $r^*(x)$ and treatment effect $\tau^*(x)$. If we replace $b_0(x)+b_1(x) < 1$ from A2 with $b_0(x) + b_1(x) \neq 1$, then $\tau^*(x)$ is identified up to sign.

- For simplicity, suppress dependence on z. By Theorem 1 and A4,

$$\tau(v_k)M\left[b_0, b_1, r(v_0)\right] = \tau(v_0)M\left[b_0, b_1, r(v_k)\right]$$

- Again suppressing dependence on z, recall from above that

$$M[b_0, b_1, r(v_k)] = \frac{1}{1 - b_1 - b_0}\left\{1 - \frac{(1-b_1)b_0}{r(v_k)} - \frac{(1-b_0)b_1}{1-r(v_k)}\right\}$$

- Evaluating the two preceding expressions at $k = 1$ and $k = 2$ gives two equations relating the identified functions r and τ to the unknown mis-classification probabilities b_0 and b_1.
- Since each equation involves v_0, we need V to take on at least three values to get as many equations as unknowns (two).
- The proof of Theorem 2 shows that the two equations admit a unique solution, so that b_0 and b_1 are identified.
- Given knowledge of b_0 and b_1, Theorem 1 allows us to solve for r^* and τ^*.
- If V only took on two values, we could not identify b_0 and b_1 without further restrictions. However if either b_0 or b_1 were known, e.g. known to be zero as in a one-sided mis-classification setting, then we could identify the model using only a binary V.

Proof of Theorem 2. For a fixed value of the covariates z, Assumption A4 ensures that there is a subset Ω_z of the support of V – a subset that may depend on z – such that for all $v, v' \in \Omega_z$ we have $b_0(v, z) = b_0(v', z)$, $b_1(v, z) = b_1(v', z)$ and $\tau^*(v, z) = \tau^*(v', z)$. Assumption

A5 ensures that Ω_z contains at least three values: v_0, v_1 and v_2. Suppress dependence on the covariates z: let $r_k = r(v_k)$ and $\tau_k = \tau(v_k)$ for $k = 0, 1, 2$. Since b_0 and b_1 do not depend on v for $v \in \Omega_z$, there is no subscript on these quantities. By Theorem 1, $\tau(v, z) = \tau^*(v, z) m(v, z)$. Again suppressing dependence on z and using the expression for m derived in the proof of Theorem 1, $\tau_k = M(b_0, b_1, r_k)\tau_k^*$ for all $k = 0, 1, 2$. But by Assumption A4, $\tau_0^* = \tau_1^* = \tau_2^*$, which yields

$$\frac{\tau_k}{M(b_0, b_1, r_k)} = \frac{\tau_\ell}{M(b_0, b_1, r_\ell)}$$

for any k, ℓ. Rearranging,

$$M(b_0, b_1, r_k)\tau_\ell = M(b_0, b_1, r_\ell)\tau_k$$

For $k \neq \ell$ this yields a nontrivial equation relating b_0 and b_1 to observables: τ_k, r_k and τ_ℓ, r_ℓ. In particular, take $\ell = 0$ and $k = 1, 2$. We obtain,

$$\begin{aligned}
0 &= M(b_0, b_1, r_k)\tau_0 - M(b_0, b_1, r_0)\tau_k \\
0 &= \frac{1}{1 - b_0 - b_1}\left[1 - \frac{(1 - b_1)b_0}{r_k} - \frac{(1 - b_0)b_1}{1 - r_k}\right]\tau_0 - \frac{1}{1 - b_0 - b_1}\left[1 - \frac{(1 - b_1)b_0}{r_0} - \frac{(1 - b_0)b_1}{1 - r_0}\right]\tau_k \\
0 &= \left[1 - \frac{(1 - b_1)b_0}{r_k} - \frac{(1 - b_0)b_1}{1 - r_k}\right]\tau_0 - \left[1 - \frac{(1 - b_1)b_0}{r_0} - \frac{(1 - b_0)b_1}{1 - r_0}\right]\tau_k \\
0 &= (1 - b_1)b_0\left(\frac{\tau_k}{r_0} - \frac{\tau_0}{r_k}\right) + (1 - b_0)b_1\left(\frac{\tau_k}{1 - r_0} - \frac{\tau_0}{1 - r_k}\right) + (\tau_0 - \tau_k) \\
0 &= (1 - b_1)b_0\left(\frac{\tau_0}{r_k} - \frac{\tau_k}{r_0}\right) + (1 - b_0)b_1\left(\frac{\tau_0}{1 - r_k} - \frac{\tau_k}{1 - r_0}\right) + (\tau_k - \tau_0)
\end{aligned}$$

This is a *linear* equation of the form

$$0 = B_0 w_{0k} + B_1 w_{1k} + w_{2k}$$

where the unknowns B_0, B_1 are defined as

$$\begin{aligned}
B_0 &= b_0(1 - b_1) \\
B_1 &= b_1(1 - b_0)
\end{aligned}$$

and the observable constants w_{0k}, w_{1k}, w_{2k} are

$$\begin{aligned}
w_{0k} &= \frac{\tau_0}{r_k} - \frac{\tau_k}{r_0} \\
w_{1k} &= \frac{\tau_0}{1 - r_k} - \frac{\tau_k}{1 - r_0} \\
w_{2k} &= \tau_k - \tau_0
\end{aligned}$$

Since we have an equation for $k = 1$ and $k = 2$, we have a linear system of k equations in two unknowns. In matrix form,

$$\begin{bmatrix} w_{01} & w_{11} \\ w_{02} & w_{12} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \begin{bmatrix} -w_{21} \\ -w_{22} \end{bmatrix}$$

In this notation, Assumption A5 is simply $w_{01}w_{12} - w_{02}w_{11} \neq 0$ which ensures the system has a unique solution.

Now, let $s = 1 - \alpha_0 - \alpha_1$. Since $B_0 = (1 - b_1)b_0$ and $B_1 = (1 - b_0)b_1$, we have

$$(s + b_0)b_0 = (1 - b_0 - b_1 + b_0)b_0 = (1 - b_1)b_0 = B_0$$

and

$$\begin{aligned} B_0 - B_1 + 1 - s &= (1 - b_1)b_0 - (1 - b_0)b_1 + 1 - (1 - b_0 - b_1) \\ &= b_0 - b_1b_0 - b_1 + b_1b_0 + b_0 + b_1 = 2b_0 \end{aligned}$$

Substituting $2b_0 = B_0 - B_1 + 1 - s$ into $B_0 = (s + b_0)b_0$, we obtain

$$\begin{aligned} 4B_0 &= (2s + 2b_0)2b_0 = (2s + B_0 - B_1 + 1 - s)(B_0 - B_1 + 1 - s) \\ &= (s + B_0 - B_1 + 1)(B_0 - B_1 + 1 - s) \\ &= [(B_0 - B_1 + 1) + s][(B_0 - B_1 + 1) - s] \\ &= (B_0 - B_1 + 1)^2 - s^2 \end{aligned}$$

and rearranging to solve for s,

$$s = \pm\sqrt{(B_0 - B_1 + 1)^2 - 4B_0}$$

This identifies s up to sign so long as $s \neq 0$, and since

$$M(b_0, b_1, r_k) = \frac{1}{1 - b_0 - b_1}\left[1 - \frac{(1 - b_1)b_0}{r_k} - \frac{(1 - b_0)b_1}{1 - r_k}\right] = \frac{1}{s}\left[1 - \frac{B_0}{r_k} - \frac{B_1}{1 - r_k}\right]$$

and $\tau_k = M(b_0, b_1, r_k)\tau^*$, it follows that τ^* is identified up to sign. If $s > 0$, then s is identified as are b_0 and b_1. To solve for the misclassification probabilities, first use the fact that $B_0 - B_1 + 1 - s = 2b_0$ as shown above to yield,

$$b_0 = (B_0 - B_1 + 1 - s)/2$$

and similarly, use the fact that

$$\begin{aligned}
-(B_0 - B_1 - 1 + s) &= -\left[(1-b_1)b_0 - (1-b_0)b_1 - b_0 - b_1\right] \\
&= -(b_0 - b_1 b_0 - b_1 + b_1 b_0 - b_0 - b_1) \\
&= 2b_1
\end{aligned}$$

we obtain

$$b_1 = -(B_0 - B_1 - 1 + s)/2.$$

Still need to verify the following:

Finally, let

$$R_k = \frac{(1-r_k^*)r_k^*}{(1-r_k)r_k}$$

Using the fact that $\tau_k = M(b_0, b_1, r_k)\tau^*$ along with

$$m = \frac{(1-b_1)r^*}{r} - \frac{b_1 r^*}{1-r}$$

the determinant condition can be expressed as

$$\left[\left(\frac{R_0}{r_1} - \frac{R_1}{r_0}\right)\left(\frac{R_0}{1-r_2} - \frac{R_2}{1-r_0}\right) - \left(\frac{R_0}{r_2} - \frac{R_2}{r_0}\right)\left(\frac{R_0}{1-r_1} - \frac{R_1}{1-r_0}\right)\right] \times (1 - b_0 - b_1)\tau^* \neq 0$$

$\square$

18 Notes on Mahajan (2006)

It should be easy to adapt the proof for Lewbel (2007) above to get the result from Mahajan (2006). We just need to replace Lewbel's assumption that $\tau^*(Z,V)$ does not depend on V with the stronger assumption that $\tau^*(Z,V)$ and $h_0^*(Z,V)$ do not depend on V. Then we only need a binary V. Does Mahajan's assumption imply that $\tau_k = \tau_\ell$? (Rather than simply that $\tau_k^* = \tau_\ell$?)

Mahajan (2006) considers regression models of the form

$$E\left[y - g(x^*, z)\right] = 0 \tag{25}$$

where x^* is an unobserved binary regressor and z is a $d_z \times 1$ vector of control regressors. Rather than x^* we observe a noisy measure x called the "surrogate" and an additional variable v that acts, in essence, as an instrumental variable. Since v does not, strictly speaking, meet the traditional requirements for an instrument, Mahajan refers to it as an "instrument-like variable" or ILV for short. Throughout the paper, Mahajan assumes that v is binary although he claims that the same idea applies to arbitrary discrete variables. The paper considers two main cases: one in which x^* is assumed to be exogenous, and another in which it is not.

18.1 The Case of Exogenous x^*

The first is based on the restriction

$$E\left[y - g(x^*, z) \mid x^*, x, z, v\right] = 0 \tag{26}$$

18.2 The Case of Endogenous x^*

While the preceding case required x^* to be exogenous, Mahajan claims (page 640) that his identification results can be extended to account for endogeneity provided that one is willing to restrict attending to additively separable models of the form

$$y = g^*(x^*, z) + \varepsilon \tag{27}$$

In this case, the ILV is assumed to satisfy the usual instrumental variables mean independence assumption

$$E\left[\varepsilon | z, v\right] = 0 \tag{28}$$

and Equation 26 is replaced by

$$E\left[y | x^*, x, z, v\right] = E\left[y | x^*, z\right] \tag{29}$$

Unfortunately, Mahajan's proof is incorrect and the model in Equation 27 is unidentified. The mistake stems from a false analogy with the identification proof in the case of exogenous x^*. In A.2 Mahajan argues, correctly, that under 27–29 knowledge of the mis-classification rates is sufficient to identify the model even when x^* is endogenous. He then appeals to Theorem 1 to argue that the mis-classification rates are indeed identified. The proof of Theorem 1, however, depends crucially on the assumption that x^* is exogenous. Without this assumption, the mis-classification rates are unidentified, as we now show For ease of exposition we consider the case without covariates. Equivalently, one can interpret all of the expressions that follow as implicitly conditioned on $z = z_a$ where z_a is a value in the support of z.[3]

Without covariates we can write

$$y = \alpha + \beta x^* + \varepsilon \tag{30}$$

where $\alpha = g^*(0)$ and $\beta = g^*(1) - g^*(0)$ and the mis-classification rates become $\eta_0 = P(x = 1|x^* = 0)$ and $\eta_1 = P(x = 0|x^* = 1)$. Now define

$$m_{jk} = E\left[\varepsilon | x^* = j, v = k\right] \tag{31}$$

19 Notation and Results for New Draft – May 8, 2017

Additively separable model

$$y = m(T^*, \mathbf{x}) + \varepsilon$$

where ε is a mean-zero error term, T^* is an endogenous binary regressor of interest and $\mathbf{x}$ is a vector of exogenous controls. Since T^* is binary, we can re-write this as linear in T^* conditional on $\mathbf{x}$

$$\begin{aligned} y &= c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon \\ \beta(\mathbf{x}) &= m(1, \mathbf{x}) - m(0, \mathbf{x}) \\ c(\mathbf{x}) &= m(0, \mathbf{x}) \end{aligned}$$

Goal is to use an instrumental variable z to identify $\beta(\mathbf{x})$ when we observe not T^* but a mis-measured binary surrogate T. Define

$$\begin{aligned} \alpha_0(\mathbf{x}, z) &= \mathbb{P}\left(T = 1 | T^* = 0, \mathbf{x}, z\right) \\ \alpha_1(\mathbf{x}, z) &= \mathbb{P}\left(T = 0 | T^* = 1, \mathbf{x}, z\right) \end{aligned}$$

Identification will only rely on two values for z so throughout the remainder of the paper we assume that z is binary and takes on values 0 and 1.

Assumption 19.1.

(i) $y = c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon$ *where* $\mathbb{E}[\varepsilon] = 0$.

(ii) $\alpha_0(\mathbf{x}, z) = \alpha_1(\mathbf{x})$, $\alpha_1(\mathbf{x}, z) = \alpha_1(\mathbf{x})$

[3]Because the covariates are held fixed throughout the proof of Mahajan's Theorem 1, there is no loss of generality.

(iii) $\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon|\mathbf{x}, z, T^*]$

(iv) $\mathbb{E}\left[T^*|\mathbf{x}, z = 0\right] \neq \mathbb{E}\left[T^*|\mathbf{x}, z = 1\right]$

(v) $\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) \neq 1$ *(T is relevant for T^*)*

(vi) $\mathbb{E}[\varepsilon|\mathbf{x}, z] = 0$

Assumption 19.2.

(i) $\mathbb{E}[\varepsilon^2|\mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon^2|\mathbf{x}, z, T^*]$

(ii) $\mathbb{E}[\varepsilon^2|\mathbf{x}, z] = \mathbb{E}[\varepsilon^2|\mathbf{x}]$

Assumption 19.3.

(i) $\mathbb{E}[\varepsilon^3|\mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon^2|\mathbf{x}, z, T^*]$

(ii) $\mathbb{E}[\varepsilon^3|\mathbf{x}, z] = \mathbb{E}[\varepsilon^3|\mathbf{x}]$

Assumption 19.4. $\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) < 1$ *(T is positively correlated with T^*)*

Notation

$$\theta_1(\mathbf{x}) = \beta(\mathbf{x})\left[1 - \{\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x})\}\right]^{-1} \tag{32}$$

$$\theta_2(\mathbf{x}) = \left[\theta_1(\mathbf{x})\right]^2\left[1 + \{\alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})\}\right] \tag{33}$$

$$\theta_3(\mathbf{x}) = \left[\theta_1(\mathbf{x})\right]^3\left[(1 - \{\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x})\})^2 + 6\alpha_0(\mathbf{x})\{1 - \alpha_1(\mathbf{x})\}\right] \tag{34}$$

$$\begin{aligned}
\mathrm{Cov}(y, z|\mathbf{x}) - \mathrm{Cov}(T, z|\mathbf{x})\theta_1(\mathbf{x}) &= 0 \\
\mathrm{Cov}(y^2, z|\mathbf{x}) - 2\mathrm{Cov}(yT, z|\mathbf{x})\theta_1(\mathbf{x}) + \mathrm{Cov}(T, z|\mathbf{x})\theta_2(\mathbf{x}) &= 0 \\
\mathrm{Cov}(y^3, z|\mathbf{x}) - 3\mathrm{Cov}(y^2T, z|\mathbf{x})\theta_1(\mathbf{x}) + 3\mathrm{Cov}(yT, z|\mathbf{x})\theta_2(\mathbf{x}) - \mathrm{Cov}(T, z|\mathbf{x})\theta_3(\mathbf{x}) &= 0
\end{aligned}$$

$$\begin{aligned}
q(\mathbf{x}) &= \mathbb{P}(z = 1|\mathbf{x}) \\
\pi(\mathbf{x}) &= \mathrm{Cov}(T, z|\mathbf{x}) \\
\eta_j(\mathbf{x}) &= \mathrm{Cov}(y^j, z|\mathbf{x}) \\
\tau_j(\mathbf{x}) &= \mathrm{Cov}(Ty^j, z|\mathbf{x})
\end{aligned}$$

$$
\begin{aligned}
\eta_1(\mathbf{x}) &= \pi(\mathbf{x})\theta_1(\mathbf{x}) \\
\eta_2(\mathbf{x}) &= 2\tau_1(\mathbf{x})\theta_1(\mathbf{x}) - \pi(\mathbf{x})\theta_2(\mathbf{x}) \\
\eta_3(\mathbf{x}) &= 3\tau_2(\mathbf{x})\theta_1(\mathbf{x}) - 3\tau_1(\mathbf{x})\theta_2(\mathbf{x}) + \pi(\mathbf{x})\theta_3(\mathbf{x})
\end{aligned}
$$

Lemma 19.1 (Various bits and pieces, e.g. η_1 equation)**.**

Lemma 19.2 (Lemma for Appendix only with Bayes' Rule)**.** *For mis-classification probabilities*

$$
\begin{aligned}
P(T^* = 1|T = 1, Z = k) &= P(T = 1|T^* = 1)\left(\frac{p_k^*}{p_k}\right) = (1-\alpha_1)\left(\frac{p_k^*}{p_k}\right) \\
P(T^* = 1|T = 0, Z = k) &= P(T = 0|T^* = 1)\left(\frac{p_k^*}{1-p_k}\right) = \alpha_1\left(\frac{p_k^*}{1-p_k}\right) \\
P(T^* = 0|T = 1, Z = k) &= P(T = 1|T^* = 0)\left(\frac{1-p_k^*}{p_k}\right) = \alpha_0\left(\frac{1-p_k^*}{p_k}\right) \\
P(T^* = 0|T = 0, Z = k) &= P(T = 0|T^* = 0)\left(\frac{1-p_k^*}{1-p_k}\right) = (1-\alpha_0)\left(\frac{1-p_k^*}{1-p_k}\right)
\end{aligned}
$$

Theorem 19.1 (Non-identification Result)**.**

Proof of Theorem 19.1. □

Proposition 19.1. *Under Assumptions ???,* $\eta_2(\mathbf{x}) = 2\tau_1(\mathbf{x})\theta_1(\mathbf{x}) - \pi(\mathbf{x})\theta_2(\mathbf{x})$.

Proof of Proposition 19.1. Since z is binary, $\text{Cov}(w, z|\mathbf{x}) = q(\mathbf{x})\,[1 - q(\mathbf{x})]\,\Delta_z\mathbb{E}(w|\mathbf{x}, z)$ for any w, where we define $\Delta_z\mathbb{E}(w|\mathbf{x}, z) \equiv \mathbb{E}(w|\mathbf{x}, z = 1) - \mathbb{E}(w|\mathbf{x}, z = 0)$. Hence it suffices to show that

$$
\Delta_z\mathbb{E}(y^2|\mathbf{x}, z) = 2\Delta_z\mathbb{E}(Ty|\mathbf{x}, z)\theta_1(\mathbf{x}) - \Delta_z\mathbb{E}(T|\mathbf{x}, z)\theta_2(\mathbf{x}).
$$

By iterated expectations

$$
\begin{aligned}
\mathbb{E}\left(y^2|\mathbf{x}, z\right) &= \mathbb{E}(y^2|\mathbf{x}, T^* = 0, z)\mathbb{P}(T^* = 0|\mathbf{x}, z) + \mathbb{E}(y^2|\mathbf{x}, T^* = 1, z)\mathbb{P}(T^* = 1|\mathbf{x}, z) \\
\mathbb{E}\left(Ty|\mathbf{x}, z\right) &= \mathbb{E}(Ty|\mathbf{x}, T^* = 0, z)\mathbb{P}(T^* = 0|\mathbf{x}, z) + \mathbb{E}(Ty|\mathbf{x}, T^* = 1, z)\mathbb{P}(T^* = 1|\mathbf{x}, z)
\end{aligned}
$$

and since $y = c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon$,

$$
\begin{aligned}
\mathbb{E}\left(y^2|\mathbf{x}, T^* = 1, z\right) &= [c(\mathbf{x}) + \beta(\mathbf{x})]^2 + 2\,[c(\mathbf{x}) + \beta(\mathbf{x})]\,\mathbb{E}\left(\varepsilon|\mathbf{x}, T^* = 1, z\right) + \mathbb{E}\left(\varepsilon^2|\mathbf{x}, T^* = 1, z\right)\\
\mathbb{E}\left(Ty|\mathbf{x}, T^* = 1, z\right) &= [c(\mathbf{x}) + \beta(\mathbf{x})]\,\mathbb{E}(T|\mathbf{x}, T^* = 1, z) + \mathbb{E}(T\varepsilon|\mathbf{x}, T^* = 1, z)\\
\mathbb{E}\left(y^2|\mathbf{x}, T^* = 0, z\right) &= c(\mathbf{x})^2 + 2c(\mathbf{x})\mathbb{E}\left(\varepsilon|\mathbf{x}, T^* = 0, z\right) + \mathbb{E}\left(\varepsilon^2|\mathbf{x}, T^* = 0, z\right)\\
\mathbb{E}\left(Ty|\mathbf{x}, T^* = 0, z\right) &= c(\mathbf{x})\,\mathbb{E}(T|\mathbf{x}, T^* = 0, z) + \mathbb{E}(T\varepsilon|\mathbf{x}, T^* = 0, z).
\end{aligned}
$$

using Assumptions ?? and 19.2. □

Proposition 19.2. *Under Assumptions ???,* $\eta_3(\mathbf{x}) = 3\tau_2(\mathbf{x})\theta_1(\mathbf{x}) - 3\tau_1(\mathbf{x})\theta_2(\mathbf{x}) + \pi(\mathbf{x})\theta_3(\mathbf{x})$

Proof of Proposition 19.2. □

Corollary 19.1. *Suppose that* $\eta_1(\mathbf{x}), \eta_2(\mathbf{x}), \eta_3(\mathbf{x}), \tau_1(\mathbf{x}), \tau_2(\mathbf{x})$ *and* $\pi(\mathbf{x})$ *are identified and that* $\pi(\mathbf{x}) \neq 0$. *Then, under the conditions of Lemmas ???,* $\theta_1(\mathbf{x}), \theta_2(\mathbf{x})$, *and* $\theta_3(\mathbf{x})$ *are identified.*

Proof of Corollary 19.1. For ease of notation we suppress dependence on $\mathbf{x}$ throughout this argument. Propositions ??? yield a triangular linear system of equations in $\theta_1, \theta_2, \theta_3$, namely

$$
\begin{aligned}
\pi\theta_1 &= \eta_1\\
2\tau_1\theta_1 - \pi\theta_2 &= \eta_2\\
3\tau_2\theta_1 - 3\tau_1\theta_2 + \pi\theta_3 &= \eta_3
\end{aligned}
$$

By inspection, the determinant of the system is $-\pi^3$ so a unique solution exists if z is correlated with T^*. In particular,

$$
\begin{aligned}
\theta_1 &= \eta_1/\pi\\
\theta_2 &= 2\tau_1\eta_1/\pi^2 - \eta_2/\pi\\
\theta_3 &= \eta_3/\pi - 3(\tau_2\eta_1 + \tau_1\eta_2)/\pi^2 + 6\tau_1\eta_1/\pi^3.
\end{aligned}
$$

□

Theorem 19.2 (Identification of β, α_0, α_1)**.**

Proof of Theorem 19.2. For ease of notation we suppress dependence on $\mathbf{x}$ throughout this argument. So long as $\beta \neq 0$, we can rearrange Equations 33 and 34 to obtain

$$A = \theta_2/\theta_1^2 = 1 + (\alpha_0 - \alpha_1) \tag{35}$$

$$B = \theta_3/\theta_1^3 = (1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1) \tag{36}$$

Equation 35 gives $(1-\alpha_1) = A-\alpha_0$. Hence $(1-\alpha_0-\alpha_1) = A-2\alpha_0$ and $\alpha_0(1-\alpha_1) = \alpha_0(A-\alpha_0)$. Substituting into Equation 36 and simplifying, $(A^2 - B) + 2A\alpha_0 - 2\alpha_0^2 = 0$. Substituting for α_0 analogously yields a quadratic in $(1 - \alpha_1)$ with *identical* coefficients. It follows that one root of $(A^2 - B) + 2Ar - 2r^2 = 0$ is α_0 and the other is $1 - \alpha_1$. Solving,

$$r = \frac{A}{2} \pm \sqrt{3A^2 - 2B} = \frac{1}{\theta_1^2}\left(\frac{\theta_2}{2} \pm \sqrt{3\theta_2^2 - 2\theta_1\theta_3}\right). \tag{37}$$

By Equations 33 and 34,

$$\begin{aligned} 3\theta_2^2 - 2\theta_1\theta_3 &= 3\left[\theta_1^2\left(1 + \alpha_0 - \alpha_1\right)\right]^2 - 2\theta_1\left\{\theta_1^3\left[(1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1)\right]\right\} \\ &= \theta_1^4\left\{3(1 + \alpha_0 - \alpha_1)^2 - 2\left[(1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1)\right]\right\}. \end{aligned}$$

Expanding the first term we find that

$$\begin{aligned} 3(1 + \alpha_0 - \alpha_1)^2 &= 3\left[1 + 2(\alpha_0 - \alpha_1) + (\alpha_0 - \alpha_1)^2\right] \\ &= 3 + 6\alpha_0 - 6\alpha_1 + 3\alpha_0^2 + 3\alpha_1^2 - 6\alpha_0\alpha_1 \end{aligned}$$

and expanding the second

$$\begin{aligned} 2\left[(1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1)\right] &= 2\left[1 - 2(\alpha_0 + \alpha_1) + (\alpha_0 + \alpha_1)^2 + 6\alpha_0 - 6\alpha_0\alpha_1\right] \\ &= 2 + 8\alpha_0 - 4\alpha_1 + 2\alpha_0^2 + 2\alpha_1^2 - 8\alpha_0\alpha_1. \end{aligned}$$

Therefore

$$\begin{aligned} 3\theta_2^2 - 2\theta_1\theta_3 &= \theta_1^4\left\{1 - 2\alpha_0 - 2\alpha_1 + \alpha_0^2 - \alpha_1^2 + 2\alpha_0\alpha_1\right\} \\ &= \theta_1^4\left[(1 - \alpha_0 - \alpha_1)^2\right] \end{aligned}$$

which is strictly greater than zero since $\theta_1 \neq 0$ and $\alpha_0 + \alpha_1 \neq 0$. It follows that both roots of the quadratic are real. Moreover, $3\theta_2^2/\theta_1^4 - 2\theta_3/\theta_1^3$ identifies $(1 - \alpha_0 - \alpha_1)^2$. Substituting into Equation 32, it follows that β is identified up to sign. If $\alpha_0 + \alpha_1 < 1$ then $\text{sign}(\beta) = \text{sign}(\theta_1)$ so that both the sign and magnitude of β are identified. If $\alpha_0 + \alpha_1 < 1$ then $1 - \alpha_1 > \alpha_0$ so $(1 - \alpha_1)$ is the larger root of $(A^2 - B) + 2Ar - 2r^2 = 0$ and α_0 is the smaller root. $\square$

20 Notes on Bugni, Canay & Shi (2017)

I've played around with this a bit for our example and I don't think it's going to work. The problematic step is the construction of the profiled test statistic. If we want, for example, to carry out inference for β, we need to estimate α_0 and α_1 under the null $\beta = \beta_0$. But if β_0 is small, the moment equalities contain almost no information about α_0 and α_1 so the optimization problem goes haywire.

The notation follows the paper nearly verbatim, although I have made a few minor changes to the notation to avoid ambiguity. Regarding the sequence κ_n, used to carry out moment selection, the paper only requires that $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$ as in Andrews & Soares. The leading example, and indeed the choice recommended by Andrews & Soares, is the BIC-type penalty $\kappa_n = \sqrt{\ln n}$ so in all the expressions below I replace κ_n with this quantity.

Setup Partially identified model defined by p moment inequalities and $k-p$ equalities:

$$
\begin{aligned}
E_F\left[m_j(W_i,\theta)\right] &\geq 0 \text{ for } j = 1, \cdots, p \\
E_F\left[m_j(W_i,\theta)\right] &= 0 \text{ for } j = p+1, \cdots, k
\end{aligned}
$$

where $W_1, \ldots, W_n \sim$ iid F. Function $m = (m_1, \ldots, m_k)$ is known and $\theta \in \Theta$ is a finite-dimensional parameter. The goal is to construct confidence sets for θ.

The Problem The existing literature tests the *joint* hypothesis $\theta = \theta_0$ and inverts this test to construct a confidence set (CS) for θ. In applied work, we often want inference for a *subset* of θ and the usual approach is to *project* a joint confidence set for θ. For example if $\theta = (\alpha, \beta, \gamma, \delta)$, a projection confidence set for α is the set of all values for which we can find *some* triple $(\beta_0, \gamma_0, \delta_0)$, allowed to depend on α, such that $\theta = (\alpha, \beta_0, \gamma_0, \delta_0)$ is not rejected. This approach is computationally intensive: even if we are only interested in one element of θ we have to invert the test over the high-dimensional *joint* parameter space. Projection inference can also have poor power compared to the joint inference upon which it is based.

This Paper The authors propose the *minimum resampling test*: a direct test of

$$
H_0\colon \lambda(\theta) = \lambda_0 \text{ vs. } H_1\colon \lambda(\theta) \neq \lambda_0
$$

where λ is a *function* of θ. The leading case is where λ is simply a sub-vector or individual coordinate of θ. The test controls size uniformly, has high power, and is in general more computationally efficient than projection inference whenever λ is of lower dimension that θ.

Decision Rule for Minimum Resampling (MR) Test The test rejects $H_0\colon \lambda(\theta) = \lambda_0$ when a *profiled test statistic* $T_n(\lambda_0)$ is large. The rejection rule is

$$\text{Reject if} \quad T_n(\lambda_0) > \widehat{c}_n^{MR}(\lambda_0, 1-\alpha)$$

where α is the significance level and $\widehat{c}_n^{MR}(\lambda_0, 1-\alpha)$ is the MR critical value, which is calculated by a bootstrap procedure described below.

Notation

$$\begin{aligned}
\bar{m}_{n,j}(\theta) &\equiv \frac{1}{n}\sum_{i=1}^{n} m_j(W_i,\theta), \text{ for } j = 1,\ldots,k \\
\widehat{\sigma}^2_{n,j} &\equiv \frac{1}{n}\left[m_j(W_i,\theta) - \bar{m}_j(W_i,\theta)\right]^2, \text{ for } j = 1,\ldots,k \\
[x]_- &\equiv \min\{x, 0\} \\
Q_n(\theta) &= \underbrace{\sum_{i=1}^{p}\left[\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\widehat{\sigma}_{n,j}(\theta)}\right]_-^2}_{\text{inequalities}} + \underbrace{\sum_{j=p+1}^{k}\left\{\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\widehat{\sigma}_{n,j}(\theta)}\right\}^2}_{\text{equalities}} \\
\Theta(\lambda_0) &\equiv \{\theta\in\Theta\colon \lambda(\theta) = \lambda_0\} \\
T_n(\lambda_0) &= \inf_{\theta\in\Theta(\lambda_0)} Q_n(\theta)
\end{aligned}$$

Calculating the MR Critical Value The MR critical value $\widehat{c}_n(\lambda_0, 1-\alpha)$ is calculated from *two different approximations* to the sampling distribution of $T_n(\lambda_0)$: *discarded resampling* (DR) and *penalized resampling* (PR), specifically

$$\begin{aligned}
\widehat{c}_n^{MR}(\lambda_0, 1-\alpha) &\equiv \text{(conditional) } 1-\alpha \text{ quantile of } T_n^{MR}(\lambda_0) \\
T_n^{MR}(\lambda_0) &\equiv \min\left\{T_n^{DR}(\lambda_0), T_n^{PR}(\lambda_0)\right\}
\end{aligned}$$

where we approximate the distribution of T_n^{MR} by simulation, as described in more detail below. Both T_n^{DR} and T_n^{PR} are constructed from $v^*_{n,j}(\theta)$, defined as:

$$\begin{gathered}
v^*_{n,j}(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\left[\frac{m_j(W_i,\theta) - \bar{m}_{n,j}(\theta)}{\widehat{\sigma}_{n,j}(\theta)}\right], \text{ for } j = 1,\ldots,p \\
\xi_1,\cdots,\xi_n \sim \text{ iid } N(0,1) \text{ independent of the data } W_1,\cdots,W_n.
\end{gathered}$$

In particular, both the T_n^{DR} and T_n^{PR} take the form

$$\inf_{\theta\in\widetilde{\Theta}}\left(\underbrace{\sum_{j=1}^{p}\left[v_{n,j}^*(\theta)+s_j(\theta)\right]_-^2}_{\text{inequalities}}+\underbrace{\sum_{j=p+1}^{k}\left\{v_{n,j}^*(\theta)+s_j(\theta)\right\}^2}_{\text{equalities}}\right)$$

where $v_{n,j}^*(\theta)$ is as defined in the preceding pair of displayed equations, $\widetilde{\Theta}$ is some set, and $s_j(\theta)$ is a *slackness function*. The difference between DR and PR is the choice of $\widetilde{\Theta}$ and $s_j(\theta)$ as explained in the following paragraphs.

Discarded Resampling Statistic: $T_n^{DR}(\lambda_0)$

$$\begin{aligned}
T_n^{DR}(\lambda_0) &\equiv \inf_{\theta\in\widehat{\Theta}_I(\lambda_0)}\left(\sum_{j=1}^{p}\left[v_{n,j}^*(\theta)+\varphi_j(\theta)\right]_-^2+\sum_{j=p+1}^{k}\left\{v_{n,j}^*(\theta)+\varphi_j(\theta)\right\}^2\right)\\
\widehat{\Theta}_I(\lambda_0) &\equiv \{\theta\in\Theta(\lambda_0)\colon Q_n(\theta)\leq T_n(\lambda_0)\}\\
\varphi_j(\theta) &= \begin{cases}\infty & \text{if} \quad \sqrt{n}\,\bar{m}_{n,j}(\theta)/[\widehat{\sigma}_{n,j}(\theta)\sqrt{\ln n}]>1 \quad \text{and } j\leq p\\ 0 & \text{if} \quad \sqrt{n}\,\bar{m}_{n,j}(\theta)/[\widehat{\sigma}_{n,j}(\theta)\sqrt{\ln n}]\leq 1 \quad \text{or } j>p\end{cases}
\end{aligned}$$

Note that $\widehat{\Theta}_I(\lambda_0)$ contains the values of θ that *minimize* $T_n(\lambda_0)=\inf_{\theta\in\Theta(\lambda_0)}Q_n(\theta)$. Note further that φ_j only affects the moment *inequalities*: it is zero for $j>p$.

Penalized Resampling Test Statistic: $T_n^{PR}(\lambda_0)$

$$\begin{aligned}
T_n^{PR}(\lambda_0) &\equiv \inf_{\theta\in\Theta(\lambda_0)}\left(\sum_{j=1}^{p}\left[v_{n,j}^*(\theta)+\ell_j(\theta)\right]_-^2+\sum_{j=p+1}^{k}\left\{v_{n,j}^*(\theta)+\ell_j(\theta)\right\}^2\right)\\
\ell_j(\theta) &= \sqrt{n}\,\bar{m}_{n,j}(\theta)/\left[\widehat{\sigma}_{n,j}(\theta)\sqrt{\ln n}\right]
\end{aligned}$$

Note that to calculate $T_n^{PR}(\lambda_0)$ we minimize over $\Theta(\lambda_0)$ – the set of all θ for which $\lambda(\theta)=\lambda_0$. Note further that $\ell_j(\theta)$ affects *both* the moment *equalities* and inequalities.

Computational Cost In practice we generate draws for the test statistics using a large number B of simulation replications. These are essentially bootstrap draws, but we bootstrap the limit distribution rather than the data. In particular, each of these bootstrap replications makes n iid standard normal draws: $\xi_1,\ldots,\xi_n$ which are used in the calculation of *both* T_n^{DR} and T_n^{PR}. (It is important to make sure that we use the same draws for both!) The computational cost of the MR test comes from the $2B+1$ optimization problems it requires us to solve. First, we need to calculate the test statistic $T_n(\lambda_0)$ based on the *actual data* – this requires us to solve one minimization problem. Then for each of the B bootstrap replications we need to calculate T_n^{DR} and T_n^{PR} – this requires us to solve two minimization problems. The authors claim that this is in general more efficient than projection inference which requires us to search over a high-dimensional space. Presumably the gains depend sensitively on how efficiently one can carry out the required minimizations.

Sub-vector Inference The leading application of the MR test is to sub-vector inference. In this case we only want to carry out a test for some subset of the coordinates of θ. Let θ_s denote this sub-vector. Then $\Theta(\lambda_0) = \{\theta \in \Theta \colon \theta_s = \lambda_0\}$. A particularly simple example is inference for a single coordinate of θ, say θ_1. In this case, the optimization problems are over all *other* coordinates of θ besides θ_1 since $\theta_1 = \lambda_0$ under the null that we are testing.

Steps to Carry out the MR Test For simplicity, in this section I assume that we wish to carry out inference on a sub-vector of θ. Let $\theta = (\beta, \gamma)$ where β and γ could be scalars or vectors. The parameters of interest are β and our null hypothesis is $\beta = \beta_0$. Let Γ denote the parameter space for γ. I write any function of θ explicitly as a function of (β, γ) so that any calculation carried out under the null is evaluated at (β_0, γ)

1. Define the following functions of γ under the null $\beta = \beta_0$

$$
\begin{aligned}
\bar{m}_n(\gamma) &= \frac{1}{n}\sum_{i=1}^{n} m(W_i, \beta_0, \gamma) \\
\widehat{D}_n(\gamma) &= \text{diag}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[m(W_i, \beta_0, \gamma) - \bar{m}_n(\gamma)\right]\left[m(W_i, \beta_0, \gamma) - \bar{m}_n(\gamma)\right]'\right\} \\
v(\gamma) &= \sqrt{n}\widehat{D}_n^{-1/2}(\gamma)\bar{m}_n(\gamma) \\
Q(\gamma) &= \sum_{j=1}^{p}\left[v_j(\gamma)\right]_-^2 + \sum_{j=p+1}^{k} v_j^2(\gamma)
\end{aligned}
$$

2. Calculate the test statistic using the observed data $\{W_1, \ldots, W_n\}$

$$
T_n = \min_{\gamma \in \Gamma} Q(\gamma)
$$

3. Calculate $\widehat{\Gamma}_I$ – the "estimated set of minimizers" of $Q(\gamma)$

$$
\widehat{\Gamma}_I = \{\gamma \in \Gamma \colon Q(\gamma) \leq T_n\}
$$

4. For each of B independent bootstrap replications, do the following:

   (i) Draw $\xi_1, \xi_2, \ldots, \xi_n \sim$ iid $N(0,1)$

   (ii) Construct $v^*(\gamma)$ – the re-centered, bootstrap version of $v(\gamma)$ – as follows

$$v^*(\gamma) = n^{-1/2}\widehat{D}_n^{-1/2}(\gamma)\sum_{i=1}^{n}\xi_i\left[m(W_i,\beta_0,\gamma) - \bar{m}_n(\gamma)\right]$$

21 May 18th, 2017

Today I figured out that the Bugni, Canay & Shi idea isn't going to work for our example since the optimization problem involved in computing the profiled test statistic is very badly behaved under the null $\beta = \beta_0$ when β_0 is small. This means that we'll have to go back to doing joint inference over $(\alpha_0, \alpha_1, \beta)$ using Andrews & Soares. Here are a few ideas:

1. I think it could be faster to use the "asymptotic version" of Andrews and Soares. The asymptotic version involves making normal draws instead of bootstrapping the sample. Is there a tradeoff in terms of accuracy? Presumably it won't matter much for reasonably large samples.

2. θ_1 is strongly identified regardless of the values of β, α_0 and α_1. This means that if we wanted to do inference for α_0, α_1 we could use a plug-in estimator of θ_1 provided that we correct the asymptotic variance matrix estimator as described in section 10.2 of Andrews and Soares.

3. Might it be interesting to look at a test for the presence of mis-classification? This would entail testing the null that $\alpha_0 = \alpha_1 = 0$. This is a null for which the moment inequalities cannot provide any information so presumably we could just use Stock and Wright's GMM-AR test. It might also be interesting to try this with Mahajan or Lewbel and show the size distortions that arise if one does not use the AR test.

4. Our implementation of Andrews and Soares should use preliminary estimators for $\boldsymbol{\kappa}$ and the parameters associated with exogenous covariates. We can use the simplest possible estimator for $\boldsymbol{\kappa}$ under the null: three sample means, e.g. $\kappa_1 = \bar{y} - \beta_0\bar{T}/(1-\alpha_0-\alpha_1)$. Similarly, for the coefficient on the exogenous covariates you just have a regression under the null in which $y - \beta_0 T/(1-\alpha_0-\alpha_1)$ replaces y.

5. Is there a clever way to use the fact that θ_1 is strongly identified to help us carry out inference for β? Specifically, suppose we construct a confidence set for α_0, α_0 by profiling out θ_1. Can we use this to back out inference for β?

22 May 18–19, 2017: More on Andrews & Soares

Recall the notation from above:

$$\mathbb{E}\left[m_j(\mathbf{w}_i,\theta_0)\right]\begin{cases}\geq 0 & \text{for } j=1,\cdots,p\\ =0 & \text{for } j=p+1,\cdots,k \text{ where } k=p+v\end{cases}$$

$$\bar{m}_n(\theta)=\begin{bmatrix}\bar{m}_{n,1}(\theta)\\ \vdots\\ \bar{m}_{n,k}(\theta)\end{bmatrix},\quad \bar{m}_{n,j}=\frac{1}{n}\sum_{i=1}^n m_j(\mathbf{w}_i,\theta) \text{ for } j=1,\cdots,k$$

Now, let $\Sigma(\theta_0)$ denote the asymptotic variance of $\sqrt{n}\ \bar{m}_n(\theta)$. We estimate this quantity using $\widehat{\Sigma}_n(\theta)$. For iid observations, as in our example, the estimator is

$$\widehat{\Sigma}_n(\theta)=\frac{1}{n}\sum_{i=1}^n\left[m(\mathbf{w}_i,\theta)-\bar{m}_n(\theta)\right]\left[m(\mathbf{w}_i,\theta)-\bar{m}_n(\theta)\right]',\quad m(\mathbf{w}_i,\theta)=\begin{bmatrix}m_1(\mathbf{w}_i,\theta)\\ \vdots\\ m_k(\mathbf{w}_i,\theta)\end{bmatrix}$$

The test statistic takes the form $T_n(\theta)=S\left(\sqrt{n}\ \bar{m}_n(\theta),\widehat{\Sigma}(\theta)\right)$ for some real-valued function S. The example we will use is S_1, defined by

$$S_1(m,\Sigma)=\sum_{j=1}^p[m_j/\sigma_j]_-^2+\sum_{j=p+1}^{p+v}(m_j/\sigma_j)^2$$

where $m=(m_1,\cdots,m_k)'$, σ_j^2 is the jth diagonal element of Σ, and

$$[x]_-=\begin{cases}x, & \text{if } x<0\\ 0, & \text{if } x\geq 0\end{cases}$$

Calculating the Critical Value for the *Asymptotic* Version of GMS Above we described the calculation of the *bootstrap* critical value for a test of $\theta=\theta_0$. We now present an alternative method that uses iid standard normal draws to "bootstrap the limit experiment."

1. Calculate $\bar{m}_n(\theta_0)$, $\widehat{\Sigma}_n(\theta_0)$, and $T_n(\theta_0)=S\left(\sqrt{n}\ \bar{m}_n(\theta_0),\widehat{\Sigma}_n(\theta_0)\right)$

2. Calculate the following quantities:

$$\begin{aligned}
\widehat{\Omega}_n(\theta_0) &= \text{Diag}^{-1/2}\left(\widehat{\Sigma}_n(\theta_0)\right)\left(\widehat{\Sigma}_n(\theta_0)\right)\text{Diag}^{-1/2}\left(\widehat{\Sigma}_n(\theta_0)\right)\\
\xi_n(\theta_0) &= \text{Diag}^{-1/2}\left(\widehat{\Sigma}_n(\theta_0)\right)\sqrt{n}\,\bar{m}_n(\theta_0)/\sqrt{\ln n}\\
\varphi\left(\xi_n(\theta_0),\widehat{\Omega}_n(\theta_0)\right) &= \begin{cases} 0, & \text{if } \xi_j \leq 1 \text{ and } j \leq p\\ \infty & \text{if } \xi_j > 1 \text{ or } j > p\end{cases}
\end{aligned}$$

3. Draw $Z_1^*, \ldots, Z_R^* \sim$ iid $N(0_k, I_k)$ for R large.

4. The critical value is the $1-\alpha$ quantile of $\left\{S\left(\widehat{\Omega}^{1/2}(\theta_0)Z_r^* + \varphi\left(\xi_n(\theta_0),\widehat{\Omega}_n(\theta_0)\right),\widehat{\Omega}_n(\theta_0)\right)\right\}_{r=1}^R$

Clearer Explanation of the Asymptotic GMS Critical Value Calculation The preceding paragraph used the explanation from Andrews & Soares but it's a little obscure. Here's a simpler and clearer explanation. Recall that the statistics from Andrews & Soares (2010) satisfy

$$\begin{aligned}
T_n &= S\left(\sqrt{n}\,\bar{m}_n(\theta),\widehat{\Sigma}(\theta)\right) = S\left(\widehat{D}_n^{-1/2}(\theta)\sqrt{n}\,\bar{m}_n(\theta),\widehat{\Omega}_n(\theta)\right)\\
\widehat{D}_n(\theta) &= \text{diag}\left(\widehat{\Sigma}\,(\theta)\right)\\
\widehat{\Omega}_n(\theta) &= \widehat{D}_n^{-1/2}(\theta)\,\widehat{\Sigma}(\theta)\,\widehat{D}_n^{-1/2}(\theta)
\end{aligned}$$

The procedure for the asymptotic version of the GMS test is as follows:

1. Calculate $\sqrt{n}\,\bar{m}_n\left(\theta_0\right)$ and $\widehat{\Sigma}(\theta_0)$ under the null hypothesis $H_0\colon \theta = \theta_0$.

2. Calculate the test statistic $T_n(\theta_0) = S_1\left(\sqrt{n}\,\bar{m}_n\left(\theta_0\right),\widehat{\Sigma}\left(\theta_0\right)\right)$.

3. Determine which inequality moment conditions are "far from binding" under $H_0\colon \theta = \theta_0$

   - Let $j \in J = \{1, \cdots, p\}$ index the inequality moment conditions.
   - Let $\widehat{\sigma}_{n,j}(\theta_0)^2$ denote the (j,j) element of $\widehat{\Sigma}(\theta_0)$
   - For each $j \in J$ calculate the "t-statistic" $t_{n,j} = \sqrt{n}\,\bar{m}_j(\theta_0)/\widehat{\sigma}_{n,j}(\theta_0)$
   - Let $\mathcal{FB}$ denote the subset of J for which $t_{n,j} > \sqrt{\log n}$. These are the inequality moment conditions that are "far from binding" under $H_0\colon \theta = \theta_0$.

4. Calculate the asymptotic critical value by simulation as follows:

   - Draw R standard normal k-vectors: $Z_1^*, \ldots, Z_R^* \sim$ iid $N(0_k, I_k)$

- Construct the k-vector $\varphi(\theta_0)$ from its coordinates φ_j as follows:

$$\varphi_j(\theta_0) = \begin{cases} \infty & j \in \mathcal{FB} \text{ and } j \leq p \\ 0 & j > p \text{ or} \end{cases}$$

- Set $M^*_{n,r}(\theta_0) = \widehat{\Omega}_n^{1/2}(\theta_0) Z^*_r$, yielding R iid $N(0_k, \widehat{\Omega}_n)$ draws.
- Calculate $T_n^{*(r)}(\theta_0) = S_1\left(M^*_{n,r}(\theta_0) + \varphi(\theta_0),\ \widehat{\Omega}_n(\theta_0)\right)$ for $r = 1, 2, \ldots, R$. Notice that adding $\varphi(\theta_0)$ is equivalent to dropping any moment conditions that we have determined are far from binding before calculating the test statistic.
- Set $\widehat{c}_n(\theta_0, 1-\delta)$ equal to the $1-\delta$ quantile of $\left\{T_n^{*(r)}(\theta_0)\right\}_{r=1}^{R}$.

5. Reject $H_0\colon \theta = \theta_0$ if $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\delta)$

6. To construct a $(1-\delta) \times 100\%$ confidence set, invert the test of $H_0\colon \theta = \theta_0$ for $\theta_0 \in \Theta$.

Preliminary Estimation of an Identified Parameter Suppose that the moment equations take the form $m_j(W_i, \theta, \tau)$ where τ is identified under the null that $\theta = \theta_0$. Let $\widehat{\tau}_n(\theta_0)$ be a consistent, asymptotically normal estimator for τ under the null that $\theta = \theta_0$. As long as $\sqrt{n}\ \bar{m}_n(\theta, \widehat{\tau}_n(\theta))$ is asymptotically normal we can plug in our estimator of τ and carry out the GMS test almost exactly as above. Only one small change is required: the variance matrix estimator $\widehat{\Sigma}_n(\theta_0)$ needs to be adjusted to take account of the fact that τ has been estimated. This is fairly straightforward using standard calculations for moment condition models. The explanation we give here follows Hall (2005) although the notation is changed to match Andrews and Soares.

Let $m(W_i, \theta, \tau)$ be a vector of moment functions that we will use to test $\theta = \theta_0$. Some of these moment functions will enter as equalities and others as inequalities but this is totally immaterial for the derivation to follow. Let $h(W_i, \theta, \tau)$ be another collection of moment functions that will be used to construct the estimator $\widehat{\tau}_n(\theta)$ but not to test $\theta = \theta_0$. Let $g(W_i, \theta, \tau) = (m', h')'$ denote the full set of moment functions. This argument will use a mean-value expansion and hence will rely on τ being on the interior of the parameter space. So the idea here is that θ contains the "problematic" parameters, those that are weakly identified or may lie on a boundary, while τ contains the "well-behaved" parameters.

Define:

$$\bar{g}_n(\theta,\tau) = \begin{bmatrix} \bar{m}_n(\theta,\tau) \\ \bar{h}_n(\theta,\tau) \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta,\tau)$$
$$G_n(\theta,\tau) = \begin{bmatrix} M_n(\theta,\tau) \\ H_n(\theta,\tau) \end{bmatrix} = n^{-1}\sum_{i=1}^{n} \partial g(W_i,\theta,\tau)/\partial\tau'$$

Throughout this argument, we will hold θ *fixed* at θ_0. Under $\theta=\theta_0$ we have the following GMM estimator for $\tau(\theta_0)$, the true value of τ assuming that $\theta=\theta_0$,

$$\widehat{\tau}_n(\theta_0) = \arg\min_{\tau\in\mathscr{T}} \bar{h}_n(\theta_0,\tau)'\,\Xi_n\,\bar{h}_n(\theta_0,\tau)$$

where $\Xi_n \to_p \Xi$. To simplify the notation, we suppress dependence on θ. Unless otherwise specified, *every function* is evaluated at $\theta=\theta_0$, e.g.

$$\widehat{\tau}_n = \widehat{\tau}_n(\theta_0), \quad \bar{g}_n(\tau) = \bar{g}_n(\theta_0,\tau), \quad G_n(\tau) = G_n(\theta_0,\tau)$$

Now, mean-value expanding $\bar{g}_n$, viewed as a function of τ *only*, around $\tau_0 \equiv \tau(\theta_0)$,

$$\bar{g}_n\left(\widehat{\tau}_n\right) = \bar{g}_n(\tau_0) + G_n(\widehat{\tau}_n,\tau_0,\lambda_n)\left(\widehat{\tau}_n - \tau_0\right) \tag{38}$$

where the ith row of $G_n(\widehat{\tau}_n,\tau_0,\lambda_n)$ equals ith row of $G_n(\widetilde{\tau}_n^{(i)})$ and

$$\widetilde{\tau}_n^{(i)} = \lambda_{n,i}\tau_0 + (1-\lambda_{n,i})\widehat{\tau}_n$$

for some $0 \le \lambda_{n,i} \le 1$. The preceding mean-value expansion is for $\bar{g}_n = (\bar{m}_n', \bar{h}_n')'$. Restricting our attention to the sub-vector $\bar{h}_n$, we have

$$\bar{h}_n(\widehat{\tau}_n) = \bar{h}_n(\tau_0) + H_n(\widehat{\tau}_n,\tau_0,\lambda_n^h)(\widehat{\tau}_n - \tau_0)$$

where λ_n^h is the sub-vector of λ_n that corresponds to H_n. Pre-multiplying by $H_n(\widehat{\tau}_n)'\Xi_n$, we have

$$H_n(\widehat{\tau}_n)'\Xi_n\bar{h}_n(\widehat{\tau}_n) = H_n(\widehat{\tau}_n)'\Xi_n\bar{h}_n(\tau_0) + H_n(\widehat{\tau}_n)'\Xi_n H_n(\widehat{\tau}_n,\tau_0,\lambda_n^h)(\widehat{\tau}_n - \tau_0).$$

But by the first-order conditions for $\widehat{\tau}_n$, we have $H_n(\widehat{\tau}_n)'\Xi_n\bar{h}_n(\widehat{\tau}_n) = 0$ and thus

$$H_n(\widehat{\tau}_n)'\Xi_n\bar{h}_n(\tau_0) + H_n(\widehat{\tau}_n)'\Xi_n H_n(\widehat{\tau}_n,\tau_0,\lambda_n^h)(\widehat{\tau}_n - \tau_0) = 0.$$

Rearranging,

$$(\widehat{\tau}_n - \tau_0) = -\left[H_n(\widehat{\tau}_n)'\Xi_n H_n(\widehat{\tau}_n, \tau_0, \lambda_n^h)\right]^{-1} H_n(\widehat{\tau}_n)'\Xi_n \bar{h}_n(\tau_0). \tag{39}$$

Now, substituting Equation 39 into Equation 38,

$$\begin{aligned}
\bar{g}_n\left(\widehat{\tau}_n\right) &= \bar{g}_n(\tau_0) + G_n(\widehat{\tau}_n, \tau_0, \lambda_n)\left(\widehat{\tau}_n - \tau_0\right) \\
&= \bar{g}_n(\tau_0) - G_n(\widehat{\tau}_n, \tau_0, \lambda_n)\left[H_n(\widehat{\tau}_n)'\Xi_n H_n(\widehat{\tau}_n, \tau_0, \lambda_n^h)\right]^{-1} H_n(\widehat{\tau}_n)'\Xi_n \bar{h}_n(\tau_0)
\end{aligned}$$

And since $\bar{g}_n = (\bar{m}_n', \bar{h}_n')$ and $G_n = (M_n', H_n')'$,

$$\begin{aligned}
\begin{bmatrix} \sqrt{n}\,\bar{m}_n(\widehat{\tau}_n) \\ \sqrt{n}\,\bar{h}_n(\widehat{\tau}_n) \end{bmatrix} &= \begin{bmatrix} \sqrt{n}\,\bar{m}_n(\tau_0) \\ \sqrt{n}\,\bar{h}_n(\tau_0) \end{bmatrix} - \begin{bmatrix} M_n(\widehat{\tau}_n, \tau_0, \lambda_n^m) \\ H_n(\widehat{\tau}_n, \tau_0, \lambda_n^h) \end{bmatrix}\left[H_n(\widehat{\tau}_n)'\Xi_n H_n(\widehat{\tau}_n, \tau_0, \lambda_n^h)\right]^{-1} H_n(\widehat{\tau}_n)'\Xi_n \left[\sqrt{n}\,\bar{h}_n(\tau_0)\right. \\
&= \begin{bmatrix} \sqrt{n}\,\bar{m}_n(\tau_0) \\ \sqrt{n}\,\bar{h}_n(\tau_0) \end{bmatrix} - \begin{bmatrix} M(\tau_0)\left\{H(\tau_0)'\Xi H(\tau_0)\right\}^{-1} H'(\tau_0)\Xi \\ H(\tau_0)\left\{H(\tau_0)'\Xi H(\tau_0)\right\}^{-1} H'(\tau_0)\Xi \end{bmatrix} \sqrt{n}\,\bar{h}_n(\tau_0) + o_p(1)
\end{aligned}$$

where $M(\tau) = E[\partial m(W_i, \theta_0, \tau)/\partial\tau']$ and $H(\tau) = E[\partial h(W_i, \theta_0, \tau)/\partial\tau']$. We are only interested in the behavior of $\sqrt{n}\, m_n(\widehat{\tau}_n)$. Restricting attention to this sub-vector,

$$\begin{aligned}
\sqrt{n}\,\bar{m}_n(\widehat{\tau}_n) &= \sqrt{n}\,\bar{m}_n(\tau_0) - M(\tau_0)\left\{H(\tau_0)'\Xi H(\tau_0)\right\}^{-1} H'(\tau_0)\Xi \times \sqrt{n}\,\bar{h}_n(\tau_0) + o_p(1) \\
&= \begin{bmatrix} \mathbf{I}_k & -M(\tau_0)\left\{H(\tau_0)'\Xi H(\tau_0)\right\}^{-1} H'(\tau_0)\Xi \end{bmatrix} \begin{bmatrix} \sqrt{n}\,\bar{m}_n(\tau_0) \\ \sqrt{n}\,\bar{h}_n(\tau_0) \end{bmatrix} + o_p(1) \\
&= \begin{bmatrix} \mathbf{I}_k & B(\tau_0) \end{bmatrix} \sqrt{n}\,\bar{g}_n(\tau_0)
\end{aligned}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix and $B(\tau_0) = M(\tau_0)\left\{H(\tau_0)'\Xi H(\tau_0)\right\}^{-1} H'(\tau_0)\Xi$. We now have all the ingredients we need to calculate our variance matrix adjustment. Restoring explicit dependence on θ_0, we have

$$\sqrt{n}\,\bar{m}_n\left(\theta_0, \widehat{\tau}_n(\theta_0)\right) = \begin{bmatrix} \mathbf{I}_k & B\left(\theta_0, \tau(\theta_0)\right) \end{bmatrix} \sqrt{n}\,\bar{g}_n\left(\theta_0, \tau(\theta_0)\right)$$

By an appropriate Central Limit Theorem, $\sqrt{n}\,\bar{g}_n\left(\theta_0, \tau(\theta_0)\right)$ is asymptotically normal with variance matrix $\mathcal{V}\left(\theta_0, \tau(\theta_0)\right)$, partitioned as follows:

$$\mathcal{V}\left(\theta_0, \tau(\theta_0)\right) = \begin{bmatrix} \mathcal{V}_{mm}\left(\theta_0, \tau(\theta_0)\right) & \mathcal{V}_{mh}\left(\theta_0, \tau(\theta_0)\right) \\ \mathcal{V}_{hm}\left(\theta_0, \tau(\theta_0)\right) & \mathcal{V}_{hh}\left(\theta_0, \tau(\theta_0)\right) \end{bmatrix}.$$

Thus, we calculate the asymptotic variance matrix $\Sigma\left(\theta_0, \widehat{\tau}_n(\theta_0)\right)$ of $\sqrt{n}\ \bar{m}_n\left(\theta_0, \widehat{\tau}_n(\theta_0)\right)$ as follows, suppressing dependence on θ and τ for simplicity:

$$\Sigma = \begin{bmatrix} \mathbf{I}_k & B \end{bmatrix} \begin{bmatrix} \mathcal{V}_{mm} & \mathcal{V}_{mh} \\ \mathcal{V}_{hm} & \mathcal{V}_{hh} \end{bmatrix} \begin{bmatrix} \mathbf{I}_k \\ B' \end{bmatrix}$$

So to estimate $\Sigma\left(\theta_0, \tau(\theta_0)\right)$ we require estimators of $B\left(\theta_0, \tau(\theta_0)\right)$ and $\mathcal{V}\left(\theta_0, \tau(\theta_0)\right)$. In our example, $\tau(\theta_0)$ is just-identified, leading to the following simplification:

$$B = -M\left(H'\Xi H\right)^{-1} H'\Xi = -M\left(H'H\right)^{-1} H' = -MH^{-1}(H')^{-1}H' = -MH^{-1}.$$

Thus, we can construct the desired variance matrix estimator $\widehat{\Sigma}_n\left(\theta_0, \widehat{\tau}_n(\theta_0)\right)$ as follows:

$$\begin{aligned}
\widehat{\mathcal{V}}_n\left(\theta_0\right) &= \frac{1}{n}\sum_{i=1}^{n}\left[g\left(W_i, \theta_0, \widehat{\tau}_n(\theta_0)\right) - \bar{g}_n\left(\theta_0, \widehat{\tau}_n(\theta_0)\right)\right]\left[g\left(W_i, \theta_0, \widehat{\tau}_n(\theta_0)\right) - \bar{g}_n\left(\theta_0, \widehat{\tau}_n(\theta_0)\right)\right]' \\
\widehat{M}_n\left(\theta_0\right) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m\left(W_i, \theta_0, \widehat{\tau}_n(\theta_0)\right)}{\partial \tau'} \\
\widehat{H}_n\left(\theta_0\right) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial h\left(W_i, \theta_0, \widehat{\tau}_n(\theta_0)\right)}{\partial \tau'} \\
\widehat{A}_n(\theta_0) &= \begin{bmatrix} \mathbf{I}_k & -\widehat{M}_n(\theta_0)\widehat{H}_n^{-1}(\theta_0) \end{bmatrix} \\
\widehat{\Sigma}_n\left(\theta_0, \widehat{\tau}_n(\theta_0)\right) &= \widehat{A}_n(\theta_0)\widehat{\mathcal{V}}_n(\theta_0)\widehat{A}_n(\theta_0)'
\end{aligned}$$

Notice that if $M = 0$, so that the m moment functions do not depend on τ, then this reduces to the expression from Andrews & Soares *without* preliminary estimation of identified parameters.

23 May 20th – A&S (2010) w/ Preliminary Estimates

Notice that in this special case we have $\theta_2 = \beta\theta_1$. Since θ_1 is strongly identified, this means that we could concentrate it out and conduct inference for β alone. But is this a good idea? It would mean that the inequalities don't give us any information. We should possibly try this out in simulations and see what happens. Note that the third moment condition simplifies in this example as well, so that $\theta_3 = \beta^2\theta_1$. We could also try using this to see how valuable the over-identifying information can be. Could be interesting to try the same thing using quantiles to get over-identifying moment equalities. Can one carry out A&S (2010) using a continuum of moment equalities and inequalities? Could at least try it out with deciles.

Simple Example: We now return to the simple example of Andrews & Soares from above, in which we assume that $\alpha_0 = 0$ and use only the "weak" bounds for α_1. The difference is that we will now estimate both intercepts: κ_1 and κ_2. The moment equalities in this case are:

$$\mathbb{E}\left[u_1(\kappa_1, \theta_1)\right] = 0, \quad \mathbb{E}\left[u_1(\kappa_1, \theta_1)z\right] = 0, \quad \mathbb{E}\left[u_2(\kappa_2, \theta_1, \theta_2)\right] = 0, \quad \mathbb{E}\left[u_2(\kappa_2, \theta_1, \theta_2)z\right] = 0$$

where

$$\begin{aligned} u_1(\kappa_1, \theta_1) &= y - \kappa_1 - \theta_1 T & \theta_1 &= \beta/(1-\alpha_1) \\ u_2(\kappa_2, \theta_1, \theta_2) &= y^2 - \kappa_2 - \theta_1 2yT + \theta_2 T & \theta_2 &= \beta^2/(1-\alpha_1) = \beta\theta_1 \end{aligned}$$

Using the notation of the preceding section, the h block of moment equalities used for preliminary estimation of $\tau = (\kappa_1, \kappa_2)$ is

$$h(W_i, \theta, \kappa) = \left[\begin{array}{l} u_{1,i}(\kappa_1, \theta_1) \\ u_{2,i}(\kappa_2, \theta_1, \theta_2) \end{array}\right] = \left[\begin{array}{l} y_i - \kappa_1 - \theta_1 T_i \\ y_i^2 - \kappa_2 - \theta_1 2y_i T_i + \theta_2 T_i \end{array}\right]$$

while the m block of moment inequalities and equalities used for inference are

$$m(W_i, \theta, \kappa) = \left[\begin{array}{l} (1-\alpha_1) - T_i(1-z_i)/(1-q) \\ (1-\alpha_1) - T_i z_i/q \\ u_{1,i}(\kappa_1, \theta_1) z_i \\ u_{2,i}(\kappa_2, \theta_1, \theta_2) z_i \end{array}\right] = \left[\begin{array}{l} (1-\alpha_1) - T_i(1-z_i)/(1-q) \\ (1-\alpha_1) - T_i z_i/q \\ (y_i - \kappa_1 - \theta_1 T_i) z_i \\ (y_i^2 - \kappa_2 - \theta_1 2y_i T_i + \theta_2 T_i) z_i \end{array}\right]$$

Under the null $(\beta, \alpha_1) = (\beta^0, \alpha_1^0)$, or equivalently $(\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$, we can estimate κ_1 and κ_2:

$$\begin{aligned} \widehat{\kappa}_1(\theta_0) &= \frac{1}{n}\sum_{i=1}^{n} y_i - \theta_1^0 T_i \\ \widehat{\kappa}_2(\theta_0) &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \theta_1^0 2y_i T_i + \theta_2^0 T_i \end{aligned}$$

Note that this estimator is just identified, so the simplification $B = -MH^{-1}$ obtains. To calculate this quantity, we first need the expected derivative matrices:

$$H = \mathbb{E}\left[\frac{\partial h}{\partial \kappa'}\right] = \left[\begin{array}{rr} -1 & 0 \\ 0 & -1 \end{array}\right] = -\mathbf{I}_2$$

and

$$M = \mathbb{E}\left[\frac{\partial m}{\partial \kappa'}\right] = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ -q & 0 \\ 0 & -q \end{array}\right] = -q \left[\begin{array}{c} \mathbf{0}_2 \\ \mathbf{I}_2 \end{array}\right]$$

which imply

$$B = -MH^{-1} = -q \left[\begin{array}{c} \mathbf{0}_2 \\ \mathbf{I}_2 \end{array}\right] \mathbf{I}_2^{-1} = -q \left[\begin{array}{c} \mathbf{0}_2 \\ \mathbf{I}_2 \end{array}\right]$$

Notice that in our example B does not depend on parameters so it does not have to be re-computed for different null hypotheses. And since $A = \left[\begin{array}{cc} I & B \end{array}\right]$, the same is true of A. In contrast, $\widehat{\mathcal{V}}_n(\theta_0)$ *does* depend on the null:

$$\widehat{\mathcal{V}}_n(\theta_0) = \frac{1}{n}\sum_{i=1}^n \left[\begin{array}{c} m(W_i, \theta_0, \widehat{\kappa}(\theta_0)) - \bar{m}_n(\theta_0, \widehat{\kappa}(\theta_0)) \\ h(W_i, \theta_0, \widehat{\kappa}(\theta_0)) - \bar{h}_n(\theta_0, \widehat{\kappa}(\theta_0)) \end{array}\right] \left[\begin{array}{c} m(W_i, \theta_0, \widehat{\kappa}) - \bar{m}_n(\theta_0, \widehat{\kappa}) \\ h(W_i, \theta_0, \widehat{\kappa}) - \bar{h}_n(\theta_0, \widehat{\kappa}) \end{array}\right]'$$

Finally, we set $\widehat{\Sigma}_n\left(\theta_0, \widehat{\kappa}(\theta_0)\right) = A\widehat{\mathcal{V}}_n A'$ and use this in the Andrews & Soares (2010) GMS test to correctly account for the fact that κ_1 and κ_2 have been estimated in a preliminary GMM step.

24 May 22nd, 2017 – Unsolved Mystery

A while back we noticed that the CDF bounds appeared to identify the true α_0 and α_1 in our normal simulation. At one point I tried to figure out if this was really true and if so why. I think I wrote out a few notes on paper but didn't make much progress. It would be good to eventually figure this out. There are a few calculations for our simulation design above on pages 44–45 of these notes. Maybe these could be used to figure it out. I think the CDF bounds may be worth coming back to since one can impose them without making our stronger assumptions about the IV. This means they could be used with Mahajan or Lewbel, for example. One needn't use all of them in practice: something like quintile bounds could be useful.

25 June 1–3, 2017 – Second Moment Bounds

Camilo triple-checked the conditional variance inequalities from our sick-instrument paper: they're definitely correct. Moreover, in spite of the fact that we defined $u = \varepsilon + c$ in that paper, they continue to hold exactly as written since c is a constant. To see why this is the case, let $u = c + \varepsilon$. Then

$$\begin{aligned}\text{Var}\,(u|T,z) &= \mathbb{E}(u^2|T,z) - [E(u|T,z)]^2\\ &= \mathbb{E}(\varepsilon^2|T,z) + 2c\mathbb{E}(\varepsilon|T,z) + c^2 - [c + \mathbb{E}(\varepsilon|T,z)]^2\\ &= \text{Var}(\varepsilon|T,z)\end{aligned}$$

Camilo then re-wrote the inequalities as follows:

$$\begin{aligned}\mathbb{E}\left[(1-p_k-\alpha_1)y^2\mathbf{1}\left\{z=k\right\}(1-T-\alpha_1) - \mathbb{P}(z=k)\left\{\alpha_1 p_k\mu_{1k} - (1-\alpha_1)(1-p_k)\mu_{0k}\right\}^2\right] > 0\\ \mathbb{E}\left[(p_k-\alpha_0)y^2\mathbf{1}\left\{z=k\right\}(T-\alpha_0) - \mathbb{P}(z=k)\left\{(1-\alpha_0)p_k\mu_{1k} - \alpha_0(1-p_k)\mu_{0k}\right\}^2\right] > 0\end{aligned}$$

where $p_k = \mathbb{P}(T=1|z=k)$ and $\mu_{tk} = \mathbb{E}(y|T=t, z=k)$. I derived one of these expressions as well (the α_1 inequality) and my derivation matches. To incorporate these inequalities into our inference procedure, we will need auxiliary moment conditions to estimate μ_{tk} and p_k. We will continue, however, to treat $\mathbb{P}(z=k)$ as fixed in repeated sampling.

We can re-write the preceding expressions in a slightly simpler form that also makes the derivatives we'll need to take to compute the covariance matrix adjustment for the GMS test much simpler. Recall that

$$\mu_{0k} = \mathbb{E}\left(y|T=0, z_k\right) = \frac{\mathbb{E}\left[y(1-T)\mathbf{1}(z=k)\right]}{(1-p_k)\mathbb{P}(z=k)}$$

$$\mu_{1k} = \mathbb{E}\left(y|T=1, z_k\right) = \frac{\mathbb{E}\left[yT\mathbf{1}(z=k)\right]}{p_k\mathbb{P}(z=k)}$$

Rearranging,

$$\begin{aligned}\mathbb{P}(z=k)(1-p_k)\mu_{0k} &= \mathbb{E}\left[y(1-T)\mathbf{1}(z=k)\right]\\ \mathbb{P}(z=k)p_k\mu_{1k} &= \mathbb{E}\left[yT\mathbf{1}(z=k)\right]\end{aligned}$$

Thus, defining

$$\begin{aligned}m_{0k} &= \mathbb{P}(z=k)(1-p_k)\mu_{0k} = \mathbb{E}\left[y(1-T)\mathbf{1}(z=k)\right]\\ m_{1k} &= \mathbb{P}(z=k)p_k\mu_{1k} = \mathbb{E}\left[yT\mathbf{1}(z=k)\right]\end{aligned}$$

by multiplying both sides by $\mathbb{P}(z = k)$ we can write the second moment inequalities as

$$\mathbb{E}\left[\mathbb{P}(z=k)(1-p_k-\alpha_1)y^2\mathbf{1}\{z=k\}(1-T-\alpha_1) - \{\alpha_1 m_{1k} - (1-\alpha_1)m_{0k}\}^2\right] > 0$$
$$\mathbb{E}\left[\mathbb{P}(z=k)(p_k-\alpha_0)y^2\mathbf{1}\{z=k\}(T-\alpha_0) - \{(1-\alpha_0)m_{1k} - \alpha_0 m_{0k}\}^2\right] > 0$$

Thus, the six quantities for which we require preliminary estimators are

$$\begin{aligned} m_{0k} &= \mathbb{E}\left[y(1-T)\mathbf{1}(z=k)\right] \\ m_{1k} &= \mathbb{E}\left[yT\mathbf{1}(z=k)\right] \\ p_k &= \mathbb{E}[T\mathbf{1}(z=k)]/\mathbb{P}(z=k) \end{aligned}$$

for $k = 0, 1$. Again, we will treat $\mathbb{P}(z = k)$ as fixed in repeated sampling. Since m_{tk} and p_k are just sample means, they should be very precisely estimated.

Are the 2nd Moment Bounds Tighter?

The second moment bounds given above are equivalent to those from the sick-instruments paper, namely

$$(p_k-\alpha_0)\left[(1-\alpha_0)p_k\sigma_{1k}^2 - \alpha_0(1-p_k)\sigma_{0k}^2\right] \geq \alpha_0(1-\alpha_0)p_k(1-p_k)(\mu_{1k}-\mu_{0k})^2 \quad (40)$$
$$(1-p_k-\alpha_1)\left[(1-\alpha_1)(1-p_k)\sigma_{0k}^2 - \alpha_1 p_k\sigma_{1k}^2\right] \geq \alpha_1(1-\alpha_1)p_k(1-p_k)(\mu_{1k}-\mu_{0k})^2 \quad (41)$$

where I have re-written $\bar{y}_{tk}$ as μ_{tk} to match the notation we use in our derivations above and allowed for the possibility that the bounds are not strict. (This is a degenerate situation in which an unobservable variance is zero, but I just want to consider all possibilities!) We will now argue that these bounds must generically be tighter than the "weak" bounds that use only p_k. This holds even if $\beta = 0$. We will proceed by analyzing the quadratic equations along which the preceding expressions hold with equality. Throughout the following argument we assume that $p_k \neq 0, 1$ and that *at least one* of $\sigma_{0k}^2, \sigma_{1k}^2$ is strictly positive.

The Bounds for α_0 Rearranging, we can write Inequality 40 as $\varphi_k(\alpha_0) \geq 0$ where

$$\begin{aligned} \varphi_k(\alpha_0) &= A_k\alpha_0^2 + B_k^0\alpha_0 + C_k^0 \\ A_k &= p_k(1-p_k)(\mu_{1k}-\mu_{0k})^2 + (1-p_k)\sigma_{0k}^2 + p_k\sigma_{1k}^2 \\ B_k^0 &= -\left[\sigma_{1k}^2 p_k(1+p_k) + p_k(1-p_k)\sigma_{0k}^2 + p_k(1-p_k)(\mu_{1k}-\mu_{0k})^2\right] \\ C_k^0 &= p_k^2\sigma_{1k}^2. \end{aligned}$$

Since we assume that $p_k \neq 0, 1$ and that at least one of $\sigma^2_{0k}, \sigma^2_{1k}$ is positive $A_k > 0$. Thus the quadratic function $\varphi_k(\alpha_0)$ opens upwards. Now, if $\alpha_0 = 0$ then the RHS of Inequality 40 becomes zero while the LHS becomes $p_k^2\sigma^2_{1k}$. Thus, Inequality 40 is always satisfied when $\alpha_0 = 0$. Similarly $\alpha_0 = 1$, the RHS is again zero while the LHS becomes $(1-p_k)^2\sigma^2_{0k}$. Thus, inequality 40 is always satisfied when $\alpha_0 = 1$. Since α_0 is a probability it follows that, so long as φ_k has two distinct roots $r_1 < r_2$, Inequality 40 is satisfied if and only if $\alpha_0 \in [0, r_1]$ or $\alpha_0 \in [r_2, 1]$. We now show that φ generically has two distinct roots, that the bound involving r_2 is extraneous, that that r_1 is generically strictly smaller than p_k so that the second moment bound is tighter than the weak bound $\alpha_0 \leq p_k$ unless $\mu_{1k} = \mu_{0k}$ in which case the weak and second moment bounds coincide.

If $\alpha_0 = p_k$, the LHS of 40 becomes zero. There are two cases. Suppose first that $\mu_{1k} \neq \mu_{0k}$. In this case, the RHS of the inequality becomes becomes $p_k^2(1-p_k)^2(\mu_{1k} - \mu_{0k})^2 > 0$ so the inequality is violated. Thus, $\mu_{1k} \neq \mu_{0k}$ implies that φ_k has two distinct roots and that p_k is strictly *between* them. Since the weak bound gives us $\alpha_0 < p_k$, the bound arising from the larger of the two roots is extraneous. Now suppose that $\mu_{1k} = \mu_{0k}$. In this case the RHS and LHS of the inequality are both zero so $\alpha_0 = p_k$ is a root of φ_k. When $\mu_{1k} = \mu_{0k}$, the coefficients of φ_k become

$$\begin{aligned}
&a = (1-p_k)\sigma^2_{0k} + p_k\sigma^2_{1k} = \sigma^2_{0k} + p_k(\sigma^2_{1k} - \sigma^2_{0k}) \\
&b = -\left[\sigma^2_{1k}p_k(1+p_k) + p_k(1-p_k)\sigma^2_{0k}\right] = -\left[p_k\sigma^2_{1k} + p_k\sigma^2_{0k} + p_k^2(\sigma^2_{1k} - \sigma^2_{0k})\right] = -p_k\left(\sigma^2_{1k} + a\right) \\
&c = p_k^2\sigma^2_{1k}.
\end{aligned}$$

Thus, we find that

$$\begin{aligned}
b^2 - 4ac &= p_k^2(\sigma^2_{1k} + a)^2 - 4ap_k^2\sigma^2_{1k} = p_k^2\left[\sigma^4_{1k} + 2a\sigma^2_{1k} + a^2 - 4a\sigma^2_{1k}\right] \\
&= p_k^2\left(\sigma^2_{1k} - a\right)^2
\end{aligned}$$

so the roots of the quadratic are

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{p_k(\sigma^2_{1k} + a) \pm p_k(\sigma^2_{1k} - a)}{2a} = \left\{\frac{p_k\sigma^2_{1k}}{a},\, p_k\right\}$$

Substituting the definition of a,

$$\frac{p_k\sigma^2_{1k}}{a} = \frac{p_k\sigma^2_{1k}}{p_k\sigma^2_{1k} + (1-p_k)\sigma^2_{0k}}$$

Rearranging, p_k is the *smaller root* when $\sigma^2_{0k} < \sigma^2_{1k}$ and the *larger root* when the inequality is reversed. When $\sigma_{0k}^2 = \sigma_{1k}^2$, the two roots are equal.

Bounds for α_1 Proceeding analogously, we can write Inequality 41 as $\psi_k(\alpha_1) \geq 0$ where

$$\begin{aligned}\psi_k(\alpha_1) &= A_k\alpha_1^2 + B_k^1\alpha_1 + C_k^1 \\ A_k &= p_k(1-p_k)(\mu_{1k}-\mu_{0k})^2 + (1-p_k)\sigma_{0k}^2 + p_k\sigma_{1k}^2 \\ B_k^1 &= -\left[p_k(1-p_k)\sigma_{1k}^2 + (1-p_k)(2-p_k)\sigma_{0k}^2 + p_k(1-p_k)(\mu_{1k}-\mu_{0k})^2\right] \\ C_k^1 &= (1-p_k)^2\sigma_{0k}^2.\end{aligned}$$

Note that the coefficient on the quadratic term is *the same* for ψ_k and φ_k. When $\mu_{1k} = \mu_{0k}$, the coefficients of ψ_k become

$$\begin{aligned}a &= (1-p_k)\sigma_{0k}^2 + p_k\sigma_{1k}^2 = \sigma_{0k}^2 + p_k(\sigma_{1k}^2 - \sigma_{0k}^2) \\ b &= -\left[p_k(1-p_k)\sigma_{1k}^2 + (1-p_k)(2-p_k)\sigma_{0k}^2\right] = -(1-p_k)\left[p_k(\sigma_{1k}^2-\sigma_{0k}^2) + 2\sigma_{0k}^2\right] = -(1-p_k)(\sigma_{0k}^2 + a) \\ c &= (1-p_k)^2\sigma_{0k}^2\end{aligned}$$

$$\begin{aligned}b^2 - 4ac &= (1-p_k)^2(\sigma_{0k}^2 + a)^2 - 4a(1-p_k)^2\sigma_{0k}^2 = (1-p_k^2)\left[(\sigma_{0k}^4 + 2a\sigma_{0k}^2 + a^2) - 4a\sigma_{0k}^2\right] \\ &= (1-p_k)^2\left(\sigma_{0k}^4 - 2a\sigma_{0k}^2 + a^2\right) = (1-p_k)^2(\sigma_{0k}^2 - a)^2\end{aligned}$$

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{(1-p_k)(\sigma_{0k}^2 + a) \pm (1-p_k)(\sigma_{0k}^2 - a)}{2a} = \left\{\frac{(1-p_k)\sigma_{0k}^2}{a},\ (1-p_k)\right\}$$

$$\frac{(1-p_k)\sigma_{0k}^2}{a} = \frac{(1-p_k)\sigma_{0k}^2}{(1-p_k)\sigma_{0k}^2 + p_k\sigma_{1k}^2}$$

Rearranging, $1-p_k$ is the *smaller root* when $\sigma_{1k}^2 < \sigma_{0k}^2$ and the *larger root* when the inequality is reversed. When $\sigma_{0k}^2 = \sigma_{1k}^2$, the two roots are equal.

What can we say about the case where $\mu_{1k} = \mu_{0k}$? Even if $\mu_{1k} = \mu_{0k}$, the second moment bounds are generically strictly better than the weak bounds. Suppose first that $\sigma_{0k}^2 \neq \sigma_{1k}^2$. If $\sigma_{0k}^2 < \sigma_{1k}^2$ then p_k is the smaller root of φ_k but $(1-p_k)$ is the *larger* root of ψ_k so the second moment bound for α_1 is strictly better than the weak bound $\alpha_1 \leq 1 - p_k$. If instead $\sigma_{1k}^2 < \sigma_{0k}^2$, then $(1-p_k)$ is the smaller root of ψ_k but p_k is the *larger* root of φ_k so the second moment bound for α_0 is strictly better than the weak bound $\alpha_0 \leq p_k$.

Only in the non-generic case where $\mu_{0k} = \mu_{1k}^2$ *and* $\sigma_{0k}^2 = \sigma_{1k}^2$ for all k do the weak bounds and second moment bounds coincide. This turns out to imply that the second moment bounds remain informative *even if the treatment effect is zero.* If $\beta = 0$ then $y = c + \varepsilon$ so $\mu_{tk} = c + \mathbb{E}(\varepsilon|T = t, z = k)$ and hence

$$0 = \mu_{1k} - \mu_{0k} = \mathbb{E}(\varepsilon|T = 1, z_k) - \mathbb{E}(\varepsilon|T = 0, z_k)$$

Thus, if $\mu_{0k} = \mu_{1k}$ then we must have $\mathbb{E}(\varepsilon|T = 1, z_k) = \mathbb{E}(\varepsilon|T = 0, z_k)$ for all k. But by iterated expectations and the assumption of non-differential measurement error we have

$$\begin{aligned}
\mathbb{E}(\varepsilon|T = 1, z_k) &= \mathbb{E}(\varepsilon|T^* = 1, z_k)\mathbb{P}(T^* = 1|T = 1, z_k) + \mathbb{E}(\varepsilon|T^* = 0, z_k)\mathbb{P}(T^* = 0|T = 1, z_k)\\
\mathbb{E}(\varepsilon|T = 0, z_k) &= \mathbb{E}(\varepsilon|T^* = 1, z_k)\mathbb{P}(T^* = 1|T = 0, z_k) + \mathbb{E}(\varepsilon|T^* = 0, z_k)\mathbb{P}(T^* = 0|T = 0, z_k)
\end{aligned}$$

Since $\mathbb{E}(\varepsilon|T = 1, z_k) = \mathbb{E}(\varepsilon|T = 0, z_k)$, these equations are a linear system of the form

$$\begin{aligned}
c &= px + (1 - p)y\\
c &= qx + (1 - q)y
\end{aligned}$$

where

$$\begin{aligned}
p &= \mathbb{P}(T^* = 1|T = 1, z_k) = (1 - \alpha_1)\frac{p_k^*}{p_k}\\
q &= \mathbb{P}(T^* = 1|T = 0, z_k) = \alpha_1\left(\frac{p_k^*}{1 - p_k}\right)\\
x &= \mathbb{E}(\varepsilon|T^* = 1, z_k)\\
y &= \mathbb{E}(\varepsilon|T^* = 0, z_k)
\end{aligned}$$

Thus, unless $p = q$, we must have $\mathbb{E}(\varepsilon|T^* = 1, z_k) = \mathbb{E}(\varepsilon|T^* = 0, z_k)$ for all k. But since z is a valid instrument, $\mathbb{E}(\varepsilon|z_k) = 0$ for all k and thus, by iterated expectations

$$\begin{aligned}
0 = \mathbb{E}(\varepsilon|z_k) &= \mathbb{E}_{T^*|z_k}\left[\mathbb{E}(\varepsilon|T^*, z_k)\right]\\
&= p_k^*\mathbb{E}(\varepsilon|T^* = 1, z_k) + (1 - p_k^*)\mathbb{E}(\varepsilon|T^* = 0, z_k)\\
&= \left[p_k^* + (1 - p_k^*)\right]\mathbb{E}(\varepsilon|T^* = 1, z_k)\\
&= \mathbb{E}(\varepsilon|T^* = 1, z_k)\\
&= \mathbb{E}(\varepsilon|T^* = 0, z_k)
\end{aligned}$$

for all k. Thus, even if $\beta = 0$ it will *still* not in general be true that $\mu_{1k} = \mu_{0k}$. For this to be the case we require the additional condition that $\mathbb{E}(\varepsilon | T_t^*, z = k) = 0$ for all t, k. In other words, we require z and T^* to be *jointly* first moment independent of ε. This implies $\mathbb{E}(\varepsilon | T^*) = 0$, i.e. that T^* is *exogenous*.

Now suppose that $\beta = 0$ *and* z, T^* are jointly mean independent of ε. The second moment bounds are *still* strictly better than the weak bounds provided that $\sigma_{0k}^2 \neq \sigma_{1k}^2$. From a proof in the appendix of the sick instruments paper,

$$
\begin{aligned}
\sigma_{1k}^2 &= \frac{(1-\alpha_1)p_k^*}{p_k} s_{1k}^{*2} + \frac{\alpha_0(1-p_k^*)}{p_k} s_{0k}^{*2} + \frac{\alpha_0(1-\alpha_1)(1-p_k)^2 \left(\mu_{1k} - \mu_{0k}\right)^2}{(p_k - \alpha_0)(1 - p_k - \alpha_1)} \\
\sigma_{0k}^2 &= \frac{\alpha_1 p_k^*}{1-p_k} s_{1k}^{*2} + \frac{(1-\alpha_0)(1-p_k^*)}{1-p_k} s_{0k}^{*2} + \frac{\alpha_1(1-\alpha_0)p_k^2 \left(\mu_{1k} - \mu_{0k}\right)^2}{(p_k - \alpha_0)(1 - p_k - \alpha_1)}
\end{aligned}
$$

where $s_{tk}^{*2} = \text{Var}(\varepsilon | T^* = t, z = k)$.[4] If $\mu_{1k} = \mu_{0k}$ this reduces to

$$
\begin{aligned}
\sigma_{1k}^2 &= \frac{(1-\alpha_1)p_k^*}{p_k} s_{1k}^{*2} + \frac{\alpha_0(1-p_k^*)}{p_k} s_{0k}^{*2} \\
\sigma_{0k}^2 &= \frac{\alpha_1 p_k^*}{1-p_k} s_{1k}^{*2} + \frac{(1-\alpha_0)(1-p_k^*)}{1-p_k} s_{0k}^{*2}
\end{aligned}
$$

In other words,

$$
\begin{aligned}
\sigma_{1k}^2 &= \mathbb{P}(T^* = 1 | T = 1, z_k) s_{1k}^{*2} + \mathbb{P}(T^* = 0 | T = 1, z_k) s_{0k}^{*2} \\
\sigma_{0k}^2 &= \mathbb{P}(T^* = 1 | T = 0, z_k) s_{1k}^{*2} + \mathbb{P}(T^* = 0 | T = 0, z_k) s_{0k}^{*2}
\end{aligned}
$$

So if $\sigma_{0k}^2 = \sigma_{1k}^2$, we have a linear system of the form

$$
\begin{aligned}
c &= px + (1-p)y \\
c &= qx + (1-q)y
\end{aligned}
$$

as above, where p and q are defined as before but $x = s_{1k}^{*2}$ and $y = s_{0k}^{*2}$. Again, unless $p = q$, we must have $x = y$ which in this case means $s_{0k}^{*2} = s_{1k}^{*2}$ for all k. Now, since

$$
s_{tk}^{*2} = \mathbb{E}(\varepsilon^2 | T^* = t, z_k) - \left[\mathbb{E}(\varepsilon | T^* = t, z_k)\right]^2
$$

under the assumption that $\mathbb{E}(\varepsilon | T^* = 0, z_k) = \mathbb{E}(\varepsilon | T^* = 1, z_k)$, we see that

$$
s_{1k}^{*2} - s_{0k}^{*2} = \mathbb{E}(\varepsilon^2 | T^* = 1, z_k) - \mathbb{E}(\varepsilon^2 | T^* = 0, z_k)
$$

[4] See the explanation above for why $\text{Var}(u | T^*, z) = \text{Var}(\varepsilon | T^*, z)$.

so that $s^{*2}_{1k} = s^{*2}_{0k}$ if and only if

$$\mathbb{E}(\varepsilon^2|T^* = 1, z_k) = \mathbb{E}(\varepsilon^2|T^* = 0, z_k)$$

Suppose this is the case. Recall that we have assumed $\mathbb{E}\left(\varepsilon^2|z_k\right) = \mathbb{E}(\varepsilon^2)$ for all k. Thus, by iterated expectations,

$$\begin{aligned}
\mathbb{E}(\varepsilon^2) &= \mathbb{E}\left(\varepsilon^2|z_k\right) = \mathbb{E}_{T^*|z_k}\left[\mathbb{E}\left(\varepsilon|T^*, z_k\right)\right] \\
&= p_k^*\mathbb{E}\left(\varepsilon^2|T^* = 1, z_k\right) + (1 - p_k^*)\mathbb{E}\left(\varepsilon^2|T^* = 0, z_k\right) \\
&= \left[p_k^* + (1 - p_k^*)\right]\mathbb{E}\left(\varepsilon^2|T^* = 1, z_k\right) \\
&= \mathbb{E}\left(\varepsilon^2|T^* = 1, z_k\right) \\
&= \mathbb{E}\left(\varepsilon^2|T^* = 0, z_k\right)
\end{aligned}$$

and thus $\mathbb{E}(\varepsilon^2|T^* = t, z = k) = \mathbb{E}(\varepsilon^2)$ for all t, k. This implies $\mathbb{E}(\varepsilon^2|T^*) = \mathbb{E}(\varepsilon^2)$.

To summarize, we have shown the following:

1. Unless $\mu_{1k} = \mu_{0k}$ *and* $\sigma^2_{1k} = \sigma^2_{0k}$ for all k, then the second moment bounds are strictly tighter for at least one of α_0 and α_1.

2. If $\beta = 0$, the second moment bounds are still generically tighter. In this case, the only way to obtain $\mu_{0k} = \mu_{1k}$ and $\sigma^2_{0k} = \sigma^2_{1k}$ for all k is if $\mathbb{E}(\varepsilon|T^*, z) = 0$ and $\mathbb{E}(\varepsilon^2|T^*, z) = \mathbb{E}(\varepsilon^2)$. These conditions in turn imply $\mathbb{E}(\varepsilon|T^*) = 0$ and $\mathbb{E}(\varepsilon^2|T^*) = \mathbb{E}(\varepsilon^2)$. So even if $\beta = 0$, the second moment bounds are tighter unless T^* is exogenous *and* ε is homoskedastic with respect to T^*.