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

Francis J. DiTraglia & Camilo García-Jimeno

April 24, 2016

1 Model and Notation

Fill in material from earlier notes so we have everything in one document!

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

Thus,

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

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

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

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

2 CDF Conditions

2.1 Notation

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^*.

2.2 Bounds on α_0, α_1

Case I: No Assumptions on Z We begin by considering the bounds that we can derive for the mis-classification error rates without imposing any conditions on Z. In other words we use only the assumption that the measurement error is non-differential and the structure of the model, namely $Y = \beta T^* + U$.

2.3 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

$$F_k(\tau) - F_\ell(\tau) = p_\ell^* \left[F_{1\ell}^*(\tau + \beta) - F_{1\ell}^*(\tau) \right] - p_k^* \left[F_{1k}^*(\tau + \beta) - F_{1k}^*(\tau) \right] \tag{2.1}$$

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 2.1, 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.

3 Special Case: $\alpha_0 = 0$

In this case the expressions from above simplify to

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

for all τ.