Missing data lecture 14: Nonidentified models for MNAR data

Identifiability of selection models

We mentioned last time that the following model:

$$y_{i1} \sim \text{Normal}(x_i^T \beta, \sigma^2)$$
$$P(m_i = 1 \mid y_{i1}) = \Phi(x_i^T \gamma + \phi_2 y_{i1})$$

was identified only by the implied joint normality of the errors in the equivalent model:

$$y_{i1} = x_i^T \beta + \sigma \epsilon_{i1}$$
$$z_i = x_i^T \gamma' + \epsilon_{i2}$$
$$(\epsilon_{i1}, \epsilon_{i2}) \sim \text{Normal}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$
$$m_i = \mathbb{1}\left(z_i > 0\right)$$

Why is this the case that identification follows from the normality assumption?

The standard procedure for estimating these models is via a two step procedure:

1. Fit a probit model to the missingness indicators to learn γ'.
2. Fit a linear regression model for y_{i1} on x_i and the conditional expectation of $z_i \mid z_i < 0$, which we get from the probit model.

Let's look at the conditional expectation of $y_{i1} \mid z_i$ given $z_i < 0$. We know the conditional expectation of $y_{i1} \mid z_i = z$:

$$x_i^T \beta + \rho\sigma(z_i - x_i^T \gamma')$$

What is the conditional expectation of $z \mid z < 0$? This is called the inverse Mills ratio:

$$\mathbb{E}\left[z \mid z < 0\right] = x_i^T \gamma' - \frac{\phi(-x_i^T \gamma')}{\Phi(-x_i^T \gamma')}$$

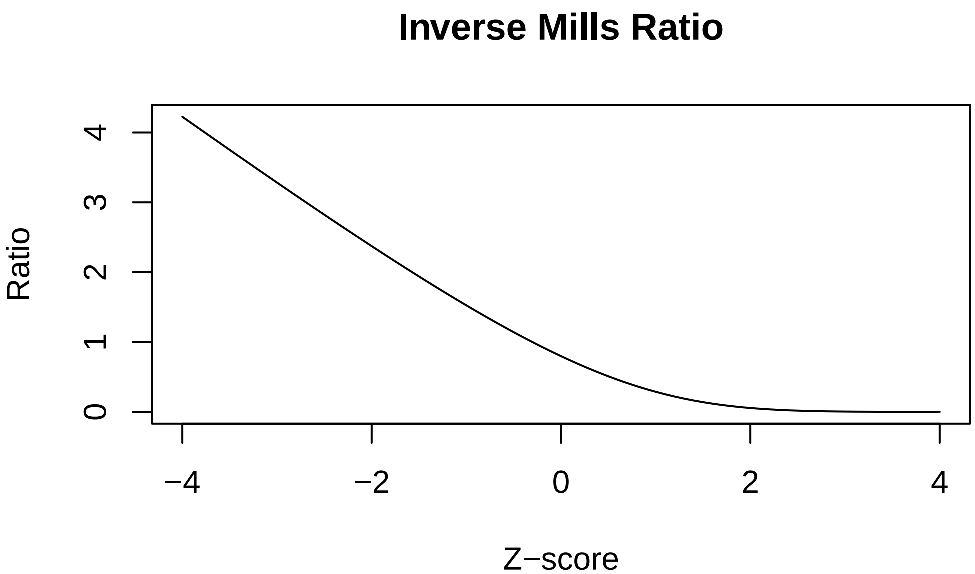
Plugging this in above gives:

$$x_i^T\beta - \rho\sigma\frac{\phi(-x_i^T\gamma')}{\Phi(-x_i^T\gamma')}$$

Thus, this is design matrix for the model we would fit to the units in which y_{i1} is observed:

$$\begin{bmatrix} x_1^T & \phi(-x_1^T\hat{\gamma}')/\Phi(-x_1^T\hat{\gamma}') \\ x_2^T & \phi(-x_2^T\hat{\gamma}')/\Phi(-x_2^T\hat{\gamma}') \\ \vdots & \vdots \\ x_n^T & \phi(-x_n^T\hat{\gamma}')/\Phi(-x_n^T\hat{\gamma}') \end{bmatrix}$$

To the extent that the function $\phi(\cdot)/\Phi(\cdot)$, which is called the inverse Mills ratio, is linear in its arguments, this extra term in the regression will be collinear with $x_i^T\beta$ and the parameter ρ won't be well-identified. Let's plot the function to see what it looks like:



Thus, the regression for y_{i1} on x_i and the Inverse Mills Ratio in the selected units is identified solely by the shape of the Inverse Mills ratio.

If we can write the model instead like:

$$y_{i1} \sim \text{Normal}(x_i^T\beta, \sigma^2)$$
$$P(m_i = 1 \mid y_{i1}) = \Phi(x_i^T\gamma + w_i^T\psi + \phi_2 y_{i1}),$$

the point estimates for β are more robust to deviations from the normality assumption. The reason for this is from above: we have another source of variation in the inverse Mills ratio, so the term won't be collinear with the predictors in the $y_{i1} \mid z_i$ regression.

There is another bivariate normal model that is identified by exclusion restrictions.

Example 1. Sensitivity analysis for pattern-mixture models Imagine we're running a household survey on income, so we can observe the address at which someone lives, but we might not learn household income due to refusals.

Let Y_{i1} be the log-household value, which we obtain from Zillow, or from property tax assessments, and Y_{i2} be the log of the response to a question about household income on a survey. Let $M_i = 1$ when Y_{i2} is missing and 0 when it is observed. Let $i = 1, \dots, r$ be the cases for which we have observed log-income, and let $i = r+1, \dots, n$ be the observations that are missing log-income. We assume that we have log-houshold value for all survey respondents.

We can use the pattern-mixture model for this scenario:

$$L(\mu, \Sigma \mid Y_{(0)} = \tilde{y}_{(0)}, M = \tilde{m}) = \prod_{i=1}^{r} (1-\omega) \mathrm{N}(y_{i1}, y_{i1} \mid, m_i = 0, \mu_0, \Sigma_0) \\ \times \prod_{i=r+1}^{n} \omega \mathrm{N}(y_{i1} \mid m_i = 1, \mu_1, \sigma_1^2)$$

We'll suppose that we can construct a variable $Y_{i2}^{\star} = Y_{i1} + \lambda Y_{i2}$, and that missingness depends only on $Y_{i2}^{\star}$: We'll suppose that the missingness mechanism has the following form:

$$P(M_i = 1 \mid Y_{i1} = y_{i1}, Y_{i2}^{\star} = y_{i2}^{\star}, \phi) = P(M_i = 1 \mid Y_{i2}^{\star} = y_{i2}^{\star}, \phi).$$

We would like to learn the following marginal expection for y_{i2}:

$$\mathbb{E}\left[Y_{i2}\right] = \mathbb{E}\left[Y_{i2} \mid M_i = 0\right](1-\omega) + \mathbb{E}\left[Y_{i2} \mid M_i = 1\right]\omega$$

where we can't calculate $\mathbb{E}\left[Y_{i2} \mid M_i = 1\right]$ directly. However, we can use the fact that the missingness mechanism does not depend on y_{i1} to identify our model.

$$\begin{aligned} f_Y(y_{i1} = y \mid Y_{i2}^{\star} = y_{i2}^{\star}, M_i = m) &= \frac{f_Y(y_{i1} = y \mid Y_{i2}^{\star} = y_{i2}^{\star}) P(M_i = m \mid y_{i1} = y, Y_{i2}^{\star} = y_{i2}^{\star})}{P(M_i = m \mid Y_{i2}^{\star} = y_{i2}^{\star})} \\ &= \frac{f_Y(y_{i1} = y \mid Y_{i2}^{\star} = y_{i2}^{\star}) P(M_i = m \mid Y_{i2}^{\star} = y_{i2}^{\star})}{P(M_i = m \mid Y_{i2}^{\star} = y_{i2}^{\star})} \\ &= f_Y(y_{i1} = y \mid Y_{i2}^{\star} = y_{i2}^{\star}) \end{aligned}$$

This means that the following holds:

$$f_Y(y_{i1} = y \mid Y_{i2}^{\star} = y_{i2}^{\star}, M_i = 1) = f_Y(y_{i1} = y \mid Y_{i2}^{\star} = y_{i2}^{\star}, M_i = 0)$$

This means that we can infer the joint distribution of Y_{i2}, Y_{i1} for $M_i = 1$ if the joint normality assumption holds. Some more algebra akin to the algebra in Lecture 13 leads to an expression of $\mu_2^{(}0)$:

$$(\hat{\mu}_2)^{(1)} = \bar{y}_2^{(0)} + \frac{\lambda s_{22} + s_{12}}{\lambda s_{12} + s_{11}} (\bar{y}_1^{(1)} - \bar{y}_1^{(0)})$$