y on m and the bias in y on x. The biases in the ML and listwise deletion results are expected because missing on y is determined by a variable not included in the model. This means that MAR does not hold. The fact that listwise deletion and ML give essentially the same results is explained by the discussion of regression in Section 10.4.2. That discussion says that individuals with missing on the dependent variable and not missing on the independent variable do not contribute to the estimation of the slope. This means that the same observations contribute to the estimation of the slope for y on m and y on x for both ML and listwise deletion.

10.5 NMAR

When the assumption of MAR is violated, the data are not missing at random (NMAR). To properly analyze NMAR data, the model needs to be expanded to include missing data indicators. Let y refer to the dependent variables of the model, m refer to the binary missing data indicators for the dependent variables, and x refer to the covariates. The full-data likelihood can be expressed in symbolic form as

$$[y, m|x] = [y|x]\,[m|y, x], \tag{10.39}$$

$$= [m|x]\,[y|m, x], \tag{10.40}$$

where (10.39) refers to the selection modeling approach (see, e.g., Diggle & Kenward, 1994) and (10.40) refers to the pattern-mixture modeling approach (see, e.g., Little, 1995). The key difference between the two approaches is that selection modeling considers m as a function of y, whereas pattern-mixture modeling considers y as a function of m. Selection modeling specifies a model for $[y|x]$, that is, the regression of y on x, and a model for $[m|y, x]$, that is, a logistic or probit regression model for the missing data indicators regressed on the dependent variables and the covariates. Pattern-mixture modeling does not model the distribution $[m|x]$, but only $[y|m, x]$, the distribution of the dependent variables regressed on the missing data indicators and the covariates.

Pattern-mixture and selection modeling are typically used when attempting to take missing data into account when data may not be MAR. Pattern-mixture modeling uses dummy covariates corresponding to the missing data patterns and lets those dummy covariates influence the parameters of the model in order to capture differences in parameter values across missing data patterns. For example, the dummy covariate for missing on the outcome in a mediation model can influence the mediator and the outcome, changing