

data on siblings to allow for family fixed effects, and to estimate within-family regressions. Fixed effects can also be associated with the villages from which survey clusters are selected, so that village means can be swept out from all the households in each cluster, thus allowing consistent estimation of the effects of quantities that vary within the village in the presence of arbitrary inter-village effects [see for example Deaton (1988) and the further discussion below].

Other studies [Rosenzweig and Wolpin (1986, 1988) and Pitt, Rosenzweig, and Gibbons (1993)] have used panel data to approach the important problem of using regression analysis to aid project evaluation. For example, Pitt et al. look at (among other things) the effects of grade-school proximity on school attendance in Indonesia combining survey with administrative data. One potential problem is that the placement of the schools is unlikely to be random — indeed the whole point of project evaluation would be to avoid random allocation — and that allocation may be influenced by unobservable local factors that themselves have a direct effect on outcomes. The simplest example would be when the government allocates schools to areas with poor attendance, so that an ultimately successful program would be one in which school attendance is the same everywhere, and where a regression analysis would show no effect of school proximity on attendance. (It is also possible that already successful areas are better at getting resources, for example through influential politicians, or by being able to turn money into votes.) Although the Indonesian data are not panels, the same administrative units (*kecamatan*s) show up in successive surveys, so that it is possible to compute a difference estimator at *kecamatan* level, a procedure that is closely related to the panel data from cross-sections methodology discussed above. This difference estimator shows much larger effects of school location on school attendance than are visible in the cross-section.

2.1.8. *Latent variables and measurement error*

Instrumental variables and panel data are only two of the possible ways of dealing with unobserved heterogeneity. In some cases, a more direct approach is available in which the data provide enough information to identify the effects of interest even in the presence of latent variables. These cases fall into the class of multiple indicator and multiple cause, or MIMIC models, which are related both to factor analysis and to models of measurement error [see in particular Goldberger (1974) and Joreskog (1973)]. Rather than discuss the general case, I look at two particular applications from the development literature.

The first is the model of imperfect fertility control of Rosenzweig and Schultz (1983), used again in Rosenzweig and Schultz (1987) and in Rosenzweig (1990), and in a somewhat different context by Pitt, Rosenzweig, and Hassan