where *educ* is years of schooling, *exper* is years of labor market experience, and *married* is a binary variable indicating marital status. The variable u, called the **error term** or **disturbance**, contains unobserved factors that affect the wage offer. Interest lies in the unknown parameters, the β_j.

We should have a concrete population in mind when specifying equation (1.1). For example, equation (1.1) could be for the population of all *working* women. In this case, it will not be difficult to obtain a random sample from the population.

All assumptions can be stated in terms of the population model. The crucial assumptions involve the relationship between u and the observable explanatory variables, *educ*, *exper*, and *married*. For example, is the expected value of u given the explanatory variables *educ*, *exper*, and *married* equal to zero? Is the variance of u conditional on the explanatory variables constant? There are reasons to think the answer to both of these questions is no, something we discuss at some length in Chapters 4 and 5. The point of raising them here is to emphasize that all such questions are most easily couched in terms of the population model.

What happens if the relevant population is *all* women over age 18? A problem arises because a random sample from this population will include women for whom the wage offer cannot be observed because they are not working. Nevertheless, we can think of a random sample being obtained, but then $wage^o$ is unobserved for women not working.

For deriving the properties of estimators, it is often useful to write the population model for a generic draw from the population. Equation (1.1) becomes

$$\log(wage_i^o) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 married_i + u_i, \tag{1.2}$$

where i indexes person. Stating assumptions in terms of u_i and $\mathbf{x}_i \equiv (educ_i, exper_i, married_i)$ is the same as stating assumptions in terms of u and $\mathbf{x}$. Throughout this book, the i subscript is reserved for indexing cross section units, such as individual, firm, city, and so on. Letters such as j, g, and h will be used to index variables, parameters, and equations.

Before ending this example, we note that using matrix notation to write equation (1.2) for all N observations adds nothing to our understanding of the model or sampling scheme; in fact, it just gets in the way because it gives the mistaken impression that the matrices tell us something about the assumptions in the underlying population. It is much better to focus on the population model (1.1).

The next example is illustrative of panel data applications.

Example 1.2 (Effect of Spillovers on Firm Output): Suppose that the population is all manufacturing firms in a country operating during a given three-year period. A