with the values of certain variables. They may be the only variables about which we have information or the only ones that we are interested in for a particular purpose. If we had more information about potential explanatory variables, we might very well specify $x_t(\boldsymbol{\beta})$ differently so as to make use of that additional information.

It is sometimes desirable to make explicit the fact that $x_t(\boldsymbol{\beta})$ represents the **conditional mean** of y_t, that is, the mean of y_t conditional on the values of a number of other variables. The set of variables on which y_t is conditioned is often referred to as an **information set**. If Ω_t denotes the information set on which the expectation of y_t is to be conditioned, one could define $x_t(\boldsymbol{\beta})$ formally as $E(y_t \,|\, \Omega_t)$. There may be more than one such information set. Thus we might well have both

$$x_{1t}(\boldsymbol{\beta}_1) \equiv E(y_t \,|\, \Omega_{1t}) \quad \text{and} \quad x_{2t}(\boldsymbol{\beta}_2) \equiv E(y_t \,|\, \Omega_{2t}),$$

where Ω_{1t} and Ω_{2t} denote two different information sets. The functions $x_{1t}(\boldsymbol{\beta}_1)$ and $x_{2t}(\boldsymbol{\beta}_2)$ might well be quite different, and we might want to estimate both of them for different purposes. There are many circumstances in which we might not want to condition on all available information. For example, if the ultimate purpose of specifying a regression function is to use it for forecasting, there may be no point in conditioning on information that will not be available at the time the forecast is to be made. Even when we do want to take account of all available information, the fact that a certain variable belongs to Ω_t does not imply that it will appear in $x_t(\boldsymbol{\beta})$, since its value may tell us nothing useful about the conditional mean of y_t, and including it may impair our ability to estimate how other variables affect that conditional mean.

For any given dependent variable y_t and information set Ω_t, one is always at liberty to consider the difference $y_t - E(y_t \,|\, \Omega_t)$ as the error term associated with the t^{th} observation. But for a *regression model* to be applicable, these differences must generally have the i.i.d. property. Actually, it is possible, when the sample size is large, to deal with cases in which the error terms are independent, but identically distributed only as regards their means, and not necessarily as regards their variances. We will discuss techniques for dealing with such cases in Chapters 16 and 17, in the latter of which we will also relax the independence assumption. As we will see in Chapter 3, however, conventional techniques for making inferences from regression models are unreliable when models lack the i.i.d. property, even when the regression function $x_t(\boldsymbol{\beta})$ is "correctly" specified. Thus we are in general not at liberty to choose an arbitrary information set and estimate a properly specified regression function based on it if we want to make inferences using conventional procedures.

There are, however, exceptional cases in which we can choose any information set we like, because models based on different information sets will always be mutually consistent. For example, suppose that the vector consisting of y_t and each of x_{1t} through x_{mt} is independently and identically