To formalize this, recall that an unbiased estimating function $D(X \mid \mu)$ that does not depend on any nuisance parameters is a k-dimensional vector function of the data X and the parameter $\mu \in R^k$ that has mean zero under all F_X in $\mathcal{M}^F$; that is, $E_{F_X} D(X \mid \mu(F_X)) = 0$. More generally, an estimating function $D(X \mid \mu, \rho)$ can also depend on a parameter ρ whose domain is the set $\mathcal{R} = \{\rho(F_X); F_X \in \mathcal{M}^F\}$ of possible values of a nuisance parameter $\rho(F_X)$. In this case, $D(X \mid \mu, \rho)$ is unbiased if

$$E_{F_X} D(X \mid \mu(F_X), \rho(F_X)) = 0 \text{ for all } F_X \in \mathcal{M}^F. \tag{1.5}$$

An estimating function $D(X \mid \mu, \rho)$ of the dimension of μ yields an estimating equation $0 = \sum_{i=1}^n D(X_i \mid \mu, \rho_n)$ for μ by replacing the parameter ρ by an estimate ρ_n and setting its empirical mean equal to zero. We can let the estimate $\rho_n(\mu)$ of $\rho(F_X)$ depend on μ; when it does, we obtain the estimating equation $0 = \sum_{i=1}^n D(X_i \mid \mu, \rho_n(\mu))$. To simplify the asymptotics of an estimator defined by a solution of an estimating equation, it is beneficial to have that the nuisance parameter ρ be locally variation-independent of μ. We have that μ and ρ are globally variation independent if each element of in $\{\mu(F_X); F_X \in M^F\} \times R$ is equal to $(\mu(F_X), \rho(F_X))$ for some $F_X \in M^F$. Similarly μ and ρ are locally variation independent if for each $F_X \in M^F$ there is a neighborhood $N(F_X) \subset M^F$ (in a natural topology) such that μ and ρ are variation independent in the local model $N(F_X)$. When μ and ρ are not variation independent we will, when possible, reparametrize the estimating function as $D^r(X \mid \mu, \rho_1) \equiv D(X \mid \mu, \rho(\mu))$, where ρ_1 and μ are now globally or locally variation independent. Although in many models it may not be possible to define a reparamaterization that yields global variation independence, one can essentially always find a reparameterization that leads to local variation independence. It is only local variation independences that is required to simplify the asymptotics of our estimators. For notational convenience, we denote the reparametrized estimating function with $D(X \mid \mu, \rho)$ again. That is, throughout the book, unless stated otherwise, one can take the parameters μ and ρ to be locally variation independent.

Let $(L_0^2(F_X), \langle f, g\rangle_{F_X} = E_{F_X} f(X)g(X))$ be the Hilbert space of mean zero one-dimensional random variables with finite variance and covariance inner product. Informally, the nuisance tangent space $T^F_{nuis}(F_X)$ at F_X is the subspace of $(L_0^2(F_X), \langle f, g\rangle_{F_X} = E_{F_X} f(X)g(X))$ defined as the closed linear span of all nuisance scores obtained by taking standard scores of one-dimensional parametric submodels that do not fluctuate the parameter of interest μ (see, e.g., Bickel, Klaassen, Ritov and Wellner, 1993). More formally, let $\{\epsilon \rightarrow F_{\epsilon,g} : g\}$ be a class of one-dimensional submodels indexed by g with parameter ϵ through F_X at $\epsilon = 0$, and let $T^F(F_X) \subset L_0^2(F_X)$ be the closure of the linear span of the corresponding scores $s(g)$ at $\epsilon = 0$. The nuisance tangent space is defined by $\left\{s(g) \in T^F(F_X) : \frac{d}{d\epsilon}\, \mu(F_{\epsilon,g})|_{\epsilon=0} = 0\right\}$; that is, these are the scores of the 1-d models that do not vary the parameter of interest μ to first order. We illustrate these concepts in the two regression model examples.

Example 1.3 (Repeated measures data with missing covariate; continuation of example 1.1) In this example, the full data structure model for $X = (Z, E, V, E^*)$ is characterized by the sole restriction (1.2). The parameter of interest is μ, and all other components of the distribution F_X represent the nuisance parameter η. The nonparametric maximum likelihood estimator of (α, η) suffers from the curse of dimensionality so that an estimating function approach to construct estimators is useful again. Lemma 2.1 in Chapter 2 proves that the orthogonal complement of the nuisance tangent space at (α, η) is given by

$$T^{F,\perp}_{nuis}(\alpha, \eta) = \{h(X^*)\epsilon(\alpha) \in L_0^2(F_X) : h(X^*) \ 1 \times p\}.$$

We will now explain the sense in which the orthogonal complement of the nuisance tangent space indeed generates all estimating functions of interest based on the full data structure X. The representation of the orthogonal complement $(\alpha, \eta) \rightarrow T^{F,\perp}_{nuis}(\alpha, \eta)$ of the nuisance tangent space as a function of (α, η) implies the following class of estimating equations for α: For any given $k \times p$ matrix function h of X^*, we could estimate α with the solution α_n of the k-dimensional estimating equation

$$0 = \frac{1}{n} \sum_{i=1}^n h(X_i^*)\epsilon_i(\alpha). \tag{1.6}$$

We will refer to h as an index of the estimating function. In other words, given a univariate class of estimating functions $(D_h : h \in \mathcal{H}^F)$ with h $1 \times p$ such that $(D_h(\cdot \mid \mu(F_X, \rho(F_X)) : h \in \mathcal{H}^F) \in T^{F,\perp}_{nuis}(F_X)$, we obtain a class of k-dimensional estimating functions $(D_h : h \in \mathcal{H}^{Fk})$ by defining for $h \in \mathcal{H}^{Fk}$, $D_h = (D_{h_1}, \ldots, D_{h_k})$. Recall that an estimator α_n is called asymptotically linear with influence curve $IC(X)$ if $\alpha_n - \alpha$ can be approximated by an empirical mean of $IC(X)$:

$$\alpha_n - \alpha = \frac{1}{n} \sum_{i=1}^n IC(X_i) + o_P(1/\sqrt{n}).$$

Under standard regularity conditions (in particular, on h), the estimator α_n solving (1.6) is asymptotically linear with influence curve

$$IC_h(X) \equiv E\left\{h(X^*)\frac{d}{d\alpha^\top} g(X^* \mid \alpha)\right\}^{-1} h(X^*)\epsilon(\alpha), \tag{1.7}$$

where $\frac{d}{d\alpha^\top} g(X^* \mid \alpha)$ is a $p \times k$ matrix and we implicitly assumed that the determinant of the $k \times k$ matrix $E\left\{h(X^*)\frac{d}{d\alpha} g(X^* \mid \alpha)\right\}$ is non zero. Thus, the influence curve at F_X is a standardized version of the estimating function itself. A well-known and important fundamental result (see e.g.,