Bickel, Klaassen, Ritov and Wellner, 1993) is that any regular asymptotically linear (RAL) estimator of μ at the full data distribution $F_{\alpha,\eta}$ has an influence curve whose components are contained in the orthogonal complement $T^{F,\perp}_{nuis}(\alpha,\eta)$ of the nuisance tangent space. We recall that an estimator μ_n of μ is regular at F_X if, for each 1-dimensional regular submodel F_ϵ with parameter ϵ and $F_{\epsilon=0} = F_X$, $\sqrt{n}(\mu_n - \mu(F_{\epsilon_n=1/\sqrt{n}})$ converges under (i.e., when sampling from) F_{ϵ_n} to a common limit distribution Z (independent of the choice of submodel). This proves that, given any regular asymptotically linear estimator of α, we can find a candidate estimating equation of the type (1.6) so that its solution is asymptotically equivalent with the estimator under appropriate regularity conditions (typically the same as needed for the given estimator). In this sense, the mapping from (α,η) to the orthogonal complement of the nuisance tangent space $T^{F,\perp}_{nuis}(\alpha,\eta)$ actually identifies *all* estimating functions for α of interest, which is a fundamental result used throughout this book.

By the multivariate central limit theorem, we have that $\sqrt{n}(\alpha_n - \alpha)$ converges in distribution to a normal limit distribution with mean vector zero and covariance matrix $\Sigma = E\{IC_h(X)IC_h(X)^\top\}$. The efficient score $h_{opt}(X^*)\,\epsilon(\alpha)$ for μ is defined to be the vector of the projections of the components of the score for μ onto the orthogonal complement $T^{F,\perp}_{nuis}$ of the nuisance tangent space. The inverse of the covariance matrix of the efficient score is referred to as the semiparametric variance bound (SVB) for the model; the asymptotic covariance matrix of every regular estimator is at least as large (in the positive definite-sense) as the SVB (Bickel, Klaassen, Ritov and Wellner, 1993). The optimal estimating function is obtained by choosing $h = h_{opt}$ so that $h_{opt}(X^*)\epsilon(\alpha)$ equals the actual efficient score. Lemma 2.1 in Chapter 2 proves that this efficient score is defined by the index

$$h_{opt}(X^*) = \frac{d}{d\alpha} g(X^* \mid \alpha)^\top_{k\times p} E(\epsilon(\alpha)\epsilon(\alpha)^\top \mid X^*)^{-1}_{p\times p}, \tag{1.8}$$

a result first proved by Chamberlain (1987). Note that this optimal index h_{opt} depends on the full data distribution F_X and is thus unknown. This suggests estimating α with the solution α_n of the estimating equation

$$0 = \frac{1}{n}\sum_{i=1}^{n} h_n(X_i^*)\epsilon_i(\alpha)$$

indexed by h_n, where h_n is an estimator of h_{opt}. Estimation of h_{opt} requires an initial estimate of α and an estimate of the $p \times p$ covariance matrix $E(\epsilon(\alpha)\epsilon(\alpha)^\top \mid X^*)$. One can obtain an initial consistent estimator $\alpha_{n,0}$ of α by solving the estimating function according to a simple choice h independent of the true parameters (e.g., $h(X^*) = d/d\alpha g(X^* \mid \alpha_0)$ at a guessed α_0). In order to construct a globally efficient estimator of α, one needs a globally consistent estimate of h_{opt}; that is, one needs a consistent estimator of the conditional covariance matrix $E(\epsilon(\alpha_{n,0})\epsilon(\alpha_{n,0})^\top \mid X^*)$ at each full data distribution in the full data model. If X^* is multidimensional, then a nonparametric estimate of the covariance matrix will require a multivariate smooth and will usually result in an estimator of α with poor performance in finite samples. As a consequence, globally efficient estimators of α are not attractive estimators in practice. However, we can obtain a so called locally consistent estimate h_n of h_{opt} by estimating the covariance matrix $E(\epsilon(\alpha_{n,0})\epsilon(\alpha_{n,0})^\top \mid X^*)$ of the estimated error term according to a lower-dimensional guessed (e.g, a multivariate normal) model. If the guessed (equivalently, working) model for the covariance matrix is correct, then the estimator α_n is an asymptotically efficient estimator of α. Furthermore, α_n remains consistent and asymptotically normal as long as h_n converges to *some* h. We say that α_n is a locally efficient estimator since it is efficient at a guessed submodel and is consistent and asymptotically normal over the whole full data structure model. The asymptotic covariance matrix of μ_n can be estimated consistently with $1/n\sum_i \widehat{IC}_{h_n}(X_i)\widehat{IC}_{h_n}(X_i)^\top$, where $\widehat{IC}_{h_n}$ estimates the influence curve (1.7). □

Example 1.4 (Repeated measures data with right-censoring; continuation of Example 1.2) Before considering the observed data structure, it is necessary to understand the estimation problem for the case where we observe n i.i.d. observations of the full data structure X. But this is is equivalent to the full data model in the previous example. Thus, for any given matrix function h of X^*, we could estimate α with the solution α_n of

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

The optimal estimating function is obtained by choosing $h = h_{opt}$, where $h_{opt}(X^*) = \{\frac{d}{d\alpha}g(X^* \mid \alpha)^\top\} E(\epsilon(\alpha)\epsilon(\alpha)^\top \mid X^*)^{-1}$. Given an estimator h_n of h_{opt}, involving an estimate of the covariance matrix $E(\epsilon(\alpha)\epsilon(\alpha)^\top \mid X^*)$ of the error term computed under a guessed lower-dimensional model, one estimates α by the solution of the corresponding estimating function

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

If the guessed model for the covariance matrix is correct, then this procedure yields a fully efficient estimator of α, and otherwise the estimator is still consistent and asymptotically normal. □

1.2.2 The curse of dimensionality in the full data model

In this subsection, we discuss the curse of dimensionality using our multivariate regression full data structure model as an illustration.