an $n \times k$ matrix, each having typical element $X_{ti}(\boldsymbol{\beta})$. These are the analogs of the vector $\boldsymbol{X}_t$ and the matrix $\boldsymbol{X}$ for the linear regression model. In the linear case, when the regression function is $\boldsymbol{X\beta}$, it is easy to see that $\boldsymbol{X}_t(\boldsymbol{\beta}) = \boldsymbol{X}_t$ and $\boldsymbol{X}(\boldsymbol{\beta}) = \boldsymbol{X}$. The big difference between the linear and nonlinear cases is that, in the latter case, $\boldsymbol{X}_t(\boldsymbol{\beta})$ and $\boldsymbol{X}(\boldsymbol{\beta})$ depend on $\boldsymbol{\beta}$.

If we multiply equation (6.10) by $n^{-1/2}$, replace $\boldsymbol{y}$ by what it is equal to under the DGP (6.01) with parameter vector $\boldsymbol{\beta}_0$, and replace $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, we obtain

$$n^{-1/2}\boldsymbol{W}^\top\big(\boldsymbol{u} + \boldsymbol{x}(\boldsymbol{\beta}_0) - \boldsymbol{x}(\hat{\boldsymbol{\beta}})\big) = \mathbf{0}. \tag{6.17}$$

The next step is to apply Taylor's Theorem at the point $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ to the components of the vector $\boldsymbol{x}(\hat{\boldsymbol{\beta}})$; see the discussion of this theorem in Section 5.6. We apply the formula (5.46) with $m = k$, replacing $f(\boldsymbol{x})$ by $x_t(\boldsymbol{\beta}_0)$ and $\boldsymbol{h}$ by the vector $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. We thus obtain, for $t = 1, \ldots, n$,

$$x_t(\hat{\boldsymbol{\beta}}) = x_t(\boldsymbol{\beta}_0) + \sum_{i=1}^{k} X_{ti}(\bar{\boldsymbol{\beta}}_t)(\hat{\beta}_i - \beta_{0i}), \tag{6.18}$$

where β_{0i} is the i^{th} element of $\boldsymbol{\beta}_0$, and the $\bar{\boldsymbol{\beta}}_t$, which play the role of $\boldsymbol{x} + \lambda\boldsymbol{h}$ in equation (5.46), satisfy the condition

$$\big\|\bar{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0\big\| \leq \big\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big\|. \tag{6.19}$$

Substituting the Taylor expansion (6.18) into (6.17) yields

$$n^{-1/2}\boldsymbol{W}^\top\boldsymbol{u} - n^{-1/2}\boldsymbol{W}^\top\boldsymbol{X}(\bar{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{0}. \tag{6.20}$$

The notation $\boldsymbol{X}(\bar{\boldsymbol{\beta}})$ is convenient, but slightly inaccurate. According to (6.18), we need different parameter vectors $\bar{\boldsymbol{\beta}}_t$ for each row of that matrix. But, since all of these vectors satisfy (6.19), it is not necessary to make this fact explicit in the notation. Thus here, and in subsequent chapters, we will refer to a vector $\bar{\boldsymbol{\beta}}$ that satisfies (6.19), without implying that it must be the *same* vector for every row of the matrix $\boldsymbol{X}(\bar{\boldsymbol{\beta}})$. This is a legitimate notational convenience, because, since $\hat{\boldsymbol{\beta}}$ is consistent, as we have seen that it is under the requirement of asymptotic identification, then so too are all of the $\bar{\boldsymbol{\beta}}_t$. Consequently, (6.20) remains true asymptotically if we replace $\bar{\boldsymbol{\beta}}$ by $\boldsymbol{\beta}_0$. Doing this, and rearranging factors of powers of n so as to work only with quantities which have suitable probability limits, yields the result that

$$n^{-1/2}\boldsymbol{W}^\top\boldsymbol{u} - n^{-1}\boldsymbol{W}^\top\boldsymbol{X}(\boldsymbol{\beta}_0)\, n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \stackrel{a}{=} \mathbf{0}. \tag{6.21}$$

This result is the starting point for all our subsequent analysis.

We need to apply a law of large numbers to the first factor of the second term of (6.21), namely, $n^{-1}\boldsymbol{W}^\top\boldsymbol{X}_0$, where for notational ease we write $\boldsymbol{X}_0 \equiv \boldsymbol{X}(\boldsymbol{\beta}_0)$.