15.2 Specification Tests Based on Artificial Regressions

In previous chapters, we have encountered numerous examples of **artificial regressions**. These include the Gauss-Newton regression (Section 6.7) and its heteroskedasticity-robust variant (Section 6.8), the OPG regression (Section 10.5), and the binary response model regression (Section 11.3). We can write any of these artificial regressions as

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{b} + \text{residuals}, \tag{15.01}$$

where $\boldsymbol{\theta}$ is a k–vector, $\boldsymbol{r}(\boldsymbol{\theta})$ is a column vector, often but by no means always of dimension equal to the sample size n, and $\boldsymbol{R}(\boldsymbol{\theta})$ is a matrix with as many rows as $\boldsymbol{r}(\boldsymbol{\theta})$ and k columns. For example, in the case of the GNR, $\boldsymbol{r}(\boldsymbol{\theta})$ is a vector of residuals, written as a function of the data and parameters, and $\boldsymbol{R}(\boldsymbol{\theta})$ is a matrix of derivatives of the regression function with respect to the parameters.

In order for (15.01) to be a valid artificial regression, the vector $\boldsymbol{r}(\boldsymbol{\theta})$ and the matrix $\boldsymbol{R}(\boldsymbol{\theta})$ must satisfy certain properties, which all of the artificial regressions we have studied do satisfy. These properties are given in outline in Exercise 8.20, and we restate them more formally here. We use a notation that was introduced in Section 9.5, whereby $\mathbb{M}$ denotes a model, μ denotes a DGP which belongs to that model, and plim_μ means a probability limit taken under the DGP μ. See the discussion in Section 9.5.

An artificial regression of the form (15.01) corresponds to a model $\mathbb{M}$ with parameter vector $\boldsymbol{\theta}$, and to a root-n consistent asymptotically normal estimator $\hat{\boldsymbol{\theta}}$ of that parameter vector, if and only if the following three conditions are satisfied:

- The artificial regressand and the artificial regressors are orthogonal when evaluated at $\hat{\boldsymbol{\theta}}$, that is,
$$\boldsymbol{R}^\top(\hat{\boldsymbol{\theta}})\boldsymbol{r}(\hat{\boldsymbol{\theta}}) = \mathbf{0}.$$
- Under any DGP $\mu \in \mathbb{M}$, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is given *either* by
$$\operatorname{Var}\big(\operatorname*{plim}_{n\to\infty}{}_\mu\, n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\mu)\big) = \operatorname*{plim}_{n\to\infty}{}_\mu \big(n^{-1}\boldsymbol{R}^\top(\acute{\boldsymbol{\theta}})\boldsymbol{R}(\acute{\boldsymbol{\theta}})\big)^{-1}, \tag{15.02}$$
where $\boldsymbol{\theta}_\mu$ is the true parameter vector for the DGP μ and n is the sample size, *or* by
$$\operatorname{Var}\big(\operatorname*{plim}_{n\to\infty}{}_\mu\, n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\mu)\big) = \operatorname*{plim}_{n\to\infty}{}_\mu\, \acute{s}^2 \big(n^{-1}\boldsymbol{R}^\top(\acute{\boldsymbol{\theta}})\boldsymbol{R}(\acute{\boldsymbol{\theta}})\big)^{-1}, \tag{15.03}$$
where $\acute{s}^2$ is the OLS estimate of the error variance obtained by running regression (15.01) with $\boldsymbol{\theta} = \acute{\boldsymbol{\theta}}$. Here, $\acute{\boldsymbol{\theta}}$ may be any root-n consistent estimator, not necessarily the same as $\hat{\boldsymbol{\theta}}$.