Definitions for Estimating Functions

Consider a model the DGPs of which are characterised, either completely or partially, by a vector $\boldsymbol{\beta}$ of parameters. Suppose that, in order to estimate $\boldsymbol{\beta}$, there are data available. For instance, for the multiple linear regression model, there would be an n-vector $\boldsymbol{y}$ with observations on the dependent variable, and an $n \times k$ matrix $\boldsymbol{X}$ with the regressors. A function $f(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta})$ of the data and the model parameters is called a **zero function** (for the given model) if, for each DGP contained in the model, μ say, characterised by parameters $\boldsymbol{\beta}(\mu)$, the expectation $\mathrm{E}\big(f(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}(\mu)\big) = 0$ when $\boldsymbol{y}$ and also possibly $\boldsymbol{X}$, are generated by the DGP μ.

Often, one can define a separate zero function for each observation in the sample. The example of this we looked at is the residual for that observation. Again, for the multiple linear regression model, the residual for observation t is $u_t(y_t, \boldsymbol{X}_t, \boldsymbol{\beta}) = y_t - \boldsymbol{X}_t\boldsymbol{\beta}$, and, if for a given DGP the true parameter vector is $\boldsymbol{\beta}_0$, this means that $y_t = \boldsymbol{X}_t\boldsymbol{\beta}_0 + u_t$, where u_t is the disturbance associated with observation t. Since u_t has expectation zero, it follows that $\mathrm{E}\big(u_t(y_t, \boldsymbol{X}_t, \boldsymbol{\beta}_0)\big) = \mathrm{E}(u_t) = 0$. In such a case, the residual $u_t(y_t, \boldsymbol{X}_t, \boldsymbol{\beta})$ is called an **elementary zero function**, the word elementary signifying that it is specific to the single observation t.

Estimation by the method of moments makes use of zero functions, usually linear combinations of the elementary zero functions of the model. The OLS estimator uses the linear combinations of the residuals defined with the regressors are coefficients, that is, the k components of the vector

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \sum_{t=1}^{n} \boldsymbol{X}_t{}^\top(y_t - \boldsymbol{X}_t\boldsymbol{\beta}) = \sum_{t=1}^{n} \boldsymbol{X}_t u_t(y_t, \boldsymbol{X}_t, \boldsymbol{\beta}).$$

Such linear combinations of the elementary zero functions are called **estimating functions**, and of course they too are zero functions.

The essence of the method of moments is that an **estimator** of the k-vector $\boldsymbol{\beta}$ can be found by solving the equations constructed by setting k suitably chosen estimating functions equal to zero. These are then the **estimating equations**, and the resulting estimator is called a **Z-estimator**, the 'Z' signifying zero.

Another huge class of estimators contains the **M-estimators**, which are defined as the parameter values that maximise or minimise (hence the 'M') a **criterion function**. The OLS estimator has traditionally been defined as an M-estimator with as criterion function the sum of squared residuals. But, as we have seen, it can also be defined as a method-of-moments estimator, and hence a Z-estimator.