be impossible to compute $\ddot{\boldsymbol{\beta}}$ from a real data set, but in the context of a Monte Carlo experiment, it is perfectly easy to do so. We know $\boldsymbol{\beta}_0$ and hence $\boldsymbol{X}_0 \equiv \boldsymbol{X}(\boldsymbol{\beta}_0)$. Using these and the error vector $\boldsymbol{u}^j$ that we generate at each replication, we can easily compute $\ddot{\boldsymbol{\beta}}^j$.

Suppose that $\theta \equiv \theta(\hat{\boldsymbol{\beta}})$ is some scalar quantity of which we wish to calculate the mean using the results of the Monte Carlo experiment. For example, if we were interested in the bias of $\hat{\beta}_2$, θ would be $\hat{\beta}_2 - \beta_{20}$; if we were interested in the mean squared error of $\hat{\beta}_3$, θ would be $(\hat{\beta}_3 - \beta_{30})^2$; if we were interested in the size of a test, θ would be 1 if the test rejected and 0 otherwise; and so on. On each replication, we obtain t_j, a realization of θ, which is equal to $\theta(\hat{\boldsymbol{\beta}}^j)$. We also obtain a control variate τ_j, which would normally be some function of $\ddot{\boldsymbol{\beta}}$. The τ_j's must be known to have mean zero and finite variance, which need not be known. If we were interested in the bias of $\hat{\beta}_2$, for example, the natural choice for τ would be $\ddot{\beta}_2$. In some other cases, it is not so obvious how to choose τ, however, and there may be several possible choices.

If the control variate τ were not available, we would estimate θ by

$$\bar{\theta} \equiv \frac{1}{N}\sum_{j=1}^{N} t_j,$$

and this naive estimator would have variance $V(\bar{\theta}) = N^{-1}V(t)$, which could be estimated by

$$\hat{V}(\bar{\theta}) = \frac{1}{N(N-1)}\sum_{j=1}^{N}\bigl(t_j - \bar{\theta}\bigr)^2.$$

When the control variate τ is available, $\bar{\theta}$ will in most cases no longer be optimal. Consider instead the control variate (CV) estimator

$$\ddot{\theta}(\lambda) \equiv \bar{\theta} - \lambda\bar{\tau}, \tag{21.10}$$

where $\bar{\tau}$ is the sample mean of the τ_j's. This estimator involves subtracting from $\bar{\theta}$ some multiple λ of the sample mean of the control variates; how λ may be chosen will be discussed in the next paragraph. On average, what is subtracted will be zero, since τ_j has population mean zero. This implies that $\ddot{\theta}(\lambda)$ must have the same population mean as $\bar{\theta}$. But, in any given sample, the mean of the τ_j's will be nonzero. If, for example, it is positive, and if τ_j and t_j are strongly positively correlated, it is very likely that $\bar{\theta}$ will also exceed its population mean. Thus, by subtracting from $\bar{\theta}$ a multiple of the mean of the τ_j's, we are likely to obtain a better estimate of θ.

The variance of the CV estimator (21.10) is

$$V\bigl(\ddot{\theta}(\lambda)\bigr) = V(\bar{\theta}) + \lambda^2 V(\bar{\tau}) - 2\lambda\,\mathrm{Cov}(\bar{\theta}, \bar{\tau}). \tag{21.11}$$