repeated samples, but it may yield results that depend on the idiosyncratic characteristics of the particular set of $\boldsymbol{X}_t$'s which was drawn.

Another possibility is to use genuine economic data for the $\boldsymbol{X}_t$'s. If these data are chosen with care, this approach can ensure that the $\boldsymbol{X}_t$'s are indeed typical of those which appear in econometric models. However, it raises a problem of how to vary the sample size. If one uses either genuine data or a single set of generated data, the matrix $n^{-1}\boldsymbol{X}^\top\boldsymbol{X}$ will change as the sample size $n$ changes. This may make it difficult to separate the effects of changes in $n$ from the effects of changes in $n^{-1}\boldsymbol{X}^\top\boldsymbol{X}$. One solution to this problem is to pick, or generate, a single set of $\boldsymbol{X}_t$'s for a sample of size $m$ and then repeat these as many times as necessary to create $\boldsymbol{X}_t$'s for samples of larger sizes. This requires that $n = cm$, where $c$ is an integer. Obvious choices for $m$ are 50 and 100; $n$ could then be any integer multiple of 50 or 100. The problem with this approach, of course, is that no matter how many replications are performed, all the results will depend on the choice of the initial set of $\boldsymbol{X}_t$'s.

In many cases, how the $\boldsymbol{X}_t$'s are chosen will not matter much. However, there are cases for which it can have a substantial impact on the results. For example, MacKinnon and White (1985) used Monte Carlo experiments to examine the finite-sample performance of various heteroskedasticity-consistent covariance matrix estimators (HCCMEs; see Section 16.3). They used 50 observations on genuine economic data for the $\boldsymbol{X}_t$'s, repeating these 50 observations as many times as necessary for each sample size. As Chesher and Jewitt (1987) subsequently showed, the performance of the estimators depends critically on the $h_t$'s, that is, the diagonal elements of the matrix $\boldsymbol{P}_X$; the larger are the largest $h_t$'s, the worse will be the finite-sample performance of tests based on all HCCMEs. When the $\boldsymbol{X}$ matrix is generated the way MacKinnon and White generated it, with $n = 50c$, all of the $h_t$'s must approach zero at a rate proportional to $1/c$ (and hence also to $1/n$). Thus MacKinnon and White were guaranteed to find that results improved rapidly as the sample size was increased. In contrast, Cragg (1983), doing Monte Carlo experiments on a related issue (see Section 17.3), generated the $\boldsymbol{X}_t$'s randomly from the lognormal distribution. This distribution has a long right-hand tail and thus occasionally throws up large values of certain $\boldsymbol{X}_t$'s. These produce relatively large values of $h_t$, and as a result the largest values of $h_t$ tend to zero at a rate very much slower than $1/n$. Thus, as the Chesher-Jewitt analysis would have predicted, Cragg found that finite-sample performance improved only slowly as the sample size was increased.

More recently, Chesher and Peters (1994) have shown that the distributions of many estimators of interest to econometricians depend crucially on the way the regressors are distributed. If the regressors are symmetrically distributed about their medians, these estimators will have special properties that do not hold in general. Since regressors used in Monte Carlo experiments might well be symmetrically distributed, there is a risk that the results of such experiments could be seriously misleading.