variance σ_0^2. We allow $\boldsymbol{X}_t$ to contain lagged dependent variables, and so we abandon the assumption of exogenous regressors and replace it with assumption (3.10) from Section 3.2, plus an analogous assumption about the variance. These two assumptions can be written as

$$\mathrm{E}(u_t \,|\, \boldsymbol{X}_t) = 0 \quad \text{and} \quad \mathrm{E}(u_t^2 \,|\, \boldsymbol{X}_t) = \sigma_0^2. \tag{4.48}$$

The first of these assumptions, which is assumption (3.10), can be referred to in two ways. From the point of view of the error terms, it says that they are **innovations**. An innovation is a random variable of which the mean is 0 conditional on the information in the explanatory variables, and so knowledge of the values taken by the latter is of no use in predicting the mean of the innovation. From the point of view of the explanatory variables $\boldsymbol{X}_t$, assumption (3.10) says that they are **predetermined** with respect to the error terms. We thus have two different ways of saying the same thing. Both can be useful, depending on the circumstances.

Although we have greatly weakened the assumptions of the classical normal linear model, we now need to make an additional assumption in order to be able to use asymptotic results. We therefore assume that the data-generating process for the explanatory variables is such that

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{S}_{\boldsymbol{X}^\top \boldsymbol{X}}, \tag{4.49}$$

where $\boldsymbol{S}_{\boldsymbol{X}^\top \boldsymbol{X}}$ is a finite, deterministic, positive definite matrix. We made this assumption previously, in Section 3.3, when we proved that the OLS estimator is consistent. Although it is often reasonable, condition (4.49) is violated in many cases. For example, it cannot hold if one of the columns of the $\boldsymbol{X}$ matrix is a linear time trend, because $\sum_{t=1}^{n} t^2$ grows at a rate faster than n.

Now consider the t statistic (4.25) for testing the hypothesis that $\beta_2 = 0$ in the model (4.21). The key to proving that (4.25), or any test statistic, has a certain **asymptotic distribution** is to write it as a function of quantities to which we can apply either a LLN or a CLT. Therefore, we rewrite (4.25) as

$$t_{\beta_2} = \left(\frac{\boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{y}}{n-k} \right)^{-1/2} \frac{n^{-1/2} \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{(n^{-1} \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2)^{1/2}}, \tag{4.50}$$

where the numerator and denominator of the second factor have both been multiplied by $n^{-1/2}$. Under the DGP (4.47), $s^2 \equiv \boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{y}/(n-k)$ tends to σ_0^2 as $n \to \infty$. This statement is equivalent to saying that the OLS error variance estimator s^2 is consistent under our weaker assumptions. Notice that s^2 is $n/(n-k)$ times the average of the $\hat{u}_t^2$. Evidently, $\operatorname{plim} n/(n-k) = 1$, and the average of the $\hat{u}_t^2$ tends to σ_0^2 by a LLN. It follows from the consistency of s^2 that the first factor in (4.50) tends to $1/\sigma_0$ as $n \to \infty$. When the data