statistic will be the same. This result assumes that we are using the efficient score form of the LM test. If we based the test on estimates of the information matrix, the two LM statistics might not be numerically the same, although they would still be the same asymptotically.

Geometrically, two different alternative hypotheses are locally equivalent if they **touch** at the null hypothesis. By this we mean not merely that the two alternative hypotheses yield the same values of their respective loglikelihood functions when restricted by the null hypothesis, as will always be the case, but also that the gradients of the two loglikelihood functions are the same, since the gradients are *tangents* to the two models that touch at the null model. In these circumstances, the two LM tests must be numerically identical.

What does it mean for two models to touch, or, to use the nongeometrical term for the property, to be locally equivalent? A circular definition would simply be that their gradients are the same at all DGPs at which the two models intersect. Statistically, it means that if one departs only slightly from the null hypothesis while respecting one of the two alternative hypotheses, then one departs from the other alternative hypothesis by an amount that is of the second order of small quantities. For instance, an AR(1) process characterized by a small autoregressive parameter ρ differs from some MA(1) process to an extent proportional only to ρ^2. To prove this formally would entail a formal definition of the distance between two DGPs, but our earlier circular definition is an operational one: If the gradient $\tilde{\boldsymbol{g}}^1$ calculated for the first alternative is the same as the gradient $\tilde{\boldsymbol{g}}^2$ for the second, then the two alternatives touch at the null. It should now be clear that this requirement is too strong: It is enough if the components of $\tilde{\boldsymbol{g}}^2$ are all linear combinations of those of $\tilde{\boldsymbol{g}}^1$ and vice versa. An example of this last possibility is provided by the local equivalence, around the null of white noise errors, of regression models with ARMA(p, q) errors on the one hand and with AR$(\max(p, q))$ errors on the other; see Section 10.8. For more examples, see Godfrey (1981) and Godfrey and Wickens (1982).

Both the geometrical and algebraic aspects of the invariance of LM tests under local equivalence are expressed by means of one simple remark: The LM test can be constructed solely on the basis of the restricted ML estimates and the *first* derivatives of the loglikelihood function evaluated at those estimates. This implies that the LM test takes no account of the curvature of the alternative hypothesis near the null.

We may summarize the results of this section as follows:

1. The LR test depends only on two maximized loglikelihood functions. It therefore cannot depend either on the parametrization of the model or on the way in which the restrictions are formulated in terms of those parameters.
2. The efficient score form of the LM test is constructed out of two ingredients, the gradient and the information matrix, which do alter under