Thus, when b_2 is a scalar, the t statistic on $\tilde{b}_2$ from the GNR (6.17) is just as valid as any of the test statistics we have been discussing.

Why does regressing residuals from the restricted model on the derivatives of $\boldsymbol{x}(\boldsymbol{\beta})$ allow us to compute valid test statistics? Why do we need to include all the derivatives and not merely those that correspond to the parameters which were restricted? The above discussion has provided formal answers to these questions, but perhaps not ones that are intuitively appealing. Let us therefore consider the matter from a slightly different point of view. In Section 5.7, we showed that Wald, LR, and LM statistics for testing the same set of restrictions are all asymptotically equal to the same random variable under the null hypothesis and that this random variable is asymptotically distributed as $\chi^2(r)$. For the nonlinear regression models we have been discussing, the LR statistic is simply the difference between $SSR(\hat{\boldsymbol{\beta}})$ and $SSR(\tilde{\boldsymbol{\beta}})$, divided by any consistent estimate of σ^2. To see why the LM statistic is valid and why the GNR must include the derivatives with respect to all parameters, we will view the LM statistic based on the GNR as a quadratic approximation to this LR statistic. That this should be the case makes sense, since the GNR itself is a linear approximation to the nonlinear regression model.

One way to view the Gauss-Newton regression is to think of it as a way of approximating the function $SSR(\boldsymbol{\beta})$ by a quadratic function that has the same first derivatives and, asymptotically, the same second derivatives at the point $\tilde{\boldsymbol{\beta}}$. This quadratic approximating function, which we will call $SSR^*(\tilde{\boldsymbol{\beta}}, \boldsymbol{b})$, is simply the sum-of-squares function for the artificial regression. It is defined as

$$SSR^*(\tilde{\boldsymbol{\beta}}, \boldsymbol{b}) = (\boldsymbol{y} - \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{X}}\boldsymbol{b})^\top(\boldsymbol{y} - \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{X}}\boldsymbol{b}).$$

The explained sum of squares from the GNR is precisely the difference between $SSR(\tilde{\boldsymbol{\beta}})$ and $SSR^*(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{b}})$. If $\tilde{\boldsymbol{\beta}}$ is reasonably close to $\hat{\boldsymbol{\beta}}$, $SSR^*(\cdot)$ should provide a good approximation to $SSR(\cdot)$ in the neighborhood of $\hat{\boldsymbol{\beta}}$. Indeed, provided that the restrictions are true and that the sample size is sufficiently large, $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ must be close to each other because they are both consistent for $\boldsymbol{\beta}_0$. Therefore, $SSR^*(\cdot)$ *must* provide a good approximation to $SSR(\cdot)$. This implies that $SSR^*(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{b}})$ will be close to $SSR(\hat{\boldsymbol{\beta}})$ and that the explained sum of squares from the GNR will provide a good approximation to $SSR(\tilde{\boldsymbol{\beta}}) - SSR(\hat{\boldsymbol{\beta}})$. When we divide the explained sum of squares by a consistent estimate of σ^2, the resulting LM test statistic should therefore be similar to the LR test statistic.

It should now be clear why the GNR has to include $\tilde{\boldsymbol{X}}_1$ as well as $\tilde{\boldsymbol{X}}_2$. If it did not, the GNR would not be minimizing $SSR^*(\tilde{\boldsymbol{\beta}}, \boldsymbol{b})$, but rather *another* approximation to $SSR(\boldsymbol{\beta})$,

$$SSR^{**}(\tilde{\boldsymbol{\beta}}, \boldsymbol{b}_2) = (\boldsymbol{y} - \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{X}}_2\boldsymbol{b}_2)^\top(\boldsymbol{y} - \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{X}}_2\boldsymbol{b}_2).$$

Although $SSR^*(\cdot)$ should normally provide a reasonably good approximation to $SSR(\cdot)$, $SSR^{**}(\cdot)$ normally will not, because it does not have enough free