region for the entire parameter vector $\boldsymbol{\beta}$, implying that $l = k$. For concreteness, we will also assume that the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ is $\hat{\boldsymbol{V}}(\hat{\boldsymbol{\beta}})$, although it could just as well be $\boldsymbol{V}_s(\hat{\boldsymbol{\beta}})$.

Let us denote the true (but unknown) value of $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_0$. Consider the quadratic form

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\top}\hat{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \tag{3.13}$$

This is just a random scalar that depends on the random vector $\hat{\boldsymbol{\beta}}$. For neither a linear nor a nonlinear regression will it actually have the χ^2 distribution with l degrees of freedom in finite samples. But it is reasonable to hope that it will be approximately distributed as $\chi^2(l)$, and in fact such an approximation is valid when the sample is large enough; see Section 5.7. Consequently, with just as much justification (or lack of it) as for the case of a single parameter, the confidence region for $\boldsymbol{\beta}$ is constructed as if (3.13) did indeed have the $\chi^2(l)$ distribution.[4]

For a given set of estimates $\hat{\boldsymbol{\beta}}$, the (approximate) confidence region at level α can be defined as the set of vectors $\boldsymbol{\beta}$ for which the value of (3.13) with $\boldsymbol{\beta}_0$ replaced by $\boldsymbol{\beta}$ is less than some critical value, say $c_\alpha(l)$. This critical value will be such that, if z is a random variable with the $\chi^2(l)$ distribution,

$$\Pr\big(z > c_\alpha(l)\big) = \alpha.$$

The confidence region is therefore the set of all $\boldsymbol{\beta}$ for which

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\top}\hat{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le c_\alpha(l). \tag{3.14}$$

Since the left-hand side of this inequality is quadratic in $\boldsymbol{\beta}$, the region is, for $l = 2$, the interior of an ellipse and, for $l > 2$, the interior of an l–dimensional ellipsoid.

Figure 3.2 illustrates what a confidence ellipse can look like in the two-parameter case. In this case, the two parameter estimates are negatively correlated, and the ellipse is centered at the parameter estimates $(\hat{\beta}_1, \hat{\beta}_2)$. Confidence intervals for β_1 and β_2 are also shown, and it should now be clear why it can be misleading to consider only these rather than the confidence ellipse. On the one hand, there are clearly many points, such as (β_1^*, β_2^*), that lie outside the confidence ellipse but inside the two confidence intervals, and on the other hand there are points, like (β_1', β_2'), that are contained in the ellipse but lie outside one or both of the confidence intervals.

[4] It is also possible, of course, to construct an approximate confidence region by using the F distribution with l and $n - k$ degrees of freedom, and this might well provide a better approximation in finite samples. Our discussion utilizes the χ^2 distribution primarily because it simplifies the exposition.