of constants, and the expected loss can be written as

$$\begin{aligned}E(y_T-\hat{y}_T)^2 &= E(\boldsymbol{\beta}'\mathbf{x}_T+u_T-\boldsymbol{\theta}'\mathbf{x}_T)^2\\ &= Eu_T^2+E\mathbf{x}_T'(\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}'-\boldsymbol{\theta}')\mathbf{x}_T\\ &= Eu_T^2+E\mathbf{x}_T'(\boldsymbol{\beta}-E\boldsymbol{\beta}+E\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}-E\boldsymbol{\beta}+E\boldsymbol{\beta}-\boldsymbol{\theta})'\mathbf{x}_T\\ &= Eu_T^2+E\left[\mathbf{x}_T'(\boldsymbol{\beta}-E\boldsymbol{\beta})(\boldsymbol{\beta}-E\boldsymbol{\beta})'\mathbf{x}_T+\mathbf{x}_T'(E\boldsymbol{\beta}-\boldsymbol{\theta})(E\boldsymbol{\beta}-\boldsymbol{\theta})'\mathbf{x}_T\right]\\ &= Eu_T^2+\operatorname{tr}S(\mathbf{x}_T)V(\boldsymbol{\beta})+(E\boldsymbol{\beta}-\boldsymbol{\theta})'S(\mathbf{x}_T)(E\boldsymbol{\beta}-\boldsymbol{\theta})\end{aligned} \tag{6.10}$$

where we have written $S(\mathbf{x}_T)=E\mathbf{x}_T\mathbf{x}_T'$. The three terms in the last line of this expression are the irreducible mean-square error Eu_T^2, a penalty for uncertainty in $\boldsymbol{\beta}$, and an additional penalty for $\boldsymbol{\theta}\neq E\boldsymbol{\beta}$, this last term wholly independent of the uncertainty in $\boldsymbol{\beta}$.

The minimal expected loss if $\boldsymbol{\theta}$ lies in the linear subspace $\mathbf{R}\boldsymbol{\theta}=\mathbf{r}$ is simply the expected posterior loss (6.10) minimized over that linear subspace. This minimization is a simple Lagrangian problem requiring the derivatives of

$$f=(E\boldsymbol{\beta}-\boldsymbol{\theta})'S(\mathbf{x}_T)(E\boldsymbol{\beta}-\boldsymbol{\theta})+2\boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\theta}-\mathbf{r})$$

to be set to zero. That is,

$$\frac{\partial f}{\partial \boldsymbol{\lambda}}=\mathbf{R}\boldsymbol{\theta}-\mathbf{r}=\mathbf{0} \tag{6.11}$$

$$\frac{\partial f}{\partial \boldsymbol{\theta}}=-S(\mathbf{x}_T)(E\boldsymbol{\beta}-\boldsymbol{\theta})+\mathbf{R}'\boldsymbol{\lambda}=\mathbf{0}. \tag{6.12}$$

These can be solved by premultiplying (6.12) by $\mathbf{R}S^{-1}(\mathbf{x}_T)$ and calculating

$$\boldsymbol{\lambda}=(\mathbf{R}S^{-1}(\mathbf{x}_T)\mathbf{R}')^{-1}(\mathbf{R}E(\boldsymbol{\beta})-\mathbf{r})$$

which can be inserted into (6.12) to obtain

$$\boldsymbol{\theta}=E(\boldsymbol{\beta})-S^{-1}(\mathbf{x}_T)\mathbf{R}'(\mathbf{R}S^{-1}(\mathbf{x}_T)\mathbf{R}')^{-1}(\mathbf{R}E(\boldsymbol{\beta})-\mathbf{r}). \tag{6.13}$$

The third term in the mean-square error (6.10) becomes

$$(E\boldsymbol{\beta}-\boldsymbol{\theta})'S(\mathbf{x}_T)(E\boldsymbol{\beta}-\boldsymbol{\theta})=(\mathbf{R}E(\boldsymbol{\beta})-\mathbf{r})'(\mathbf{R}S^{-1}(\mathbf{x}_T)\mathbf{R}')^{-1}(\mathbf{R}E(\boldsymbol{\beta})-\mathbf{r}). \tag{6.14}$$

It is obvious from the positive definiteness of the third term in (6.10) that minimal expected posterior loss requires $\boldsymbol{\theta}=E\boldsymbol{\beta}$, or by (6.14) that a restriction increases expected loss unless $\mathbf{R}E\boldsymbol{\beta}=\mathbf{r}$. A simplification thus necessarily decreases expected prediction accuracy. We assume that a simplification has benefits also, and in the absence of any clear quantitative statement of those benefits, a reasonable number to report is the percentage increase in the expected posterior loss due to the restriction $\mathbf{R}\boldsymbol{\theta}=\mathbf{r}$:

$$L^2(\mathbf{R},\mathbf{r})=\frac{(\mathbf{R}E(\boldsymbol{\beta})-\mathbf{r})'(\mathbf{R}S^{-1}(\mathbf{x}_T)\mathbf{R}')^{-1}(\mathbf{R}E(\boldsymbol{\beta})-\mathbf{r})}{Eu_T^2+\operatorname{tr}S(\mathbf{x}_T)V(\boldsymbol{\beta})}. \tag{6.15}$$

With suitable definitions of prior vagueness we have simply the least-squares results (remembering that the expected value operator is conditional on $\mathbf{X}$ and $\mathbf{Y}$)

$$E(\boldsymbol{\beta}|\mathbf{Y},\mathbf{X})=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$V(\boldsymbol{\beta}|\mathbf{Y},\mathbf{X})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Further, if the explanatory variables are independent observations from a multivariate process, we would have the $\mathbf{x}_T$ moment matrix be approximately (see Section 3.4)

$$S(\mathbf{x}_T)=E(\mathbf{x}_T\mathbf{x}_T')=\frac{\mathbf{X}'\mathbf{X}}{T}.$$

Using these in (6.13), $\boldsymbol{\theta}$ is seen to be simply the constrained least-squares estimate subject to $\mathbf{R}\boldsymbol{\beta}=\mathbf{r}$. Inserting them into (6.14), we obtain the increase in the posterior expected loss to be T^{-1} times a factor that is well known to be the increase in the error-sum squares due to the restriction. The summary L^2 becomes

$$L^2(\mathbf{R},\mathbf{r})=\frac{T^{-1}\Delta ESS}{\sigma^2\left(1+\frac{k}{T}\right)} \tag{6.16}$$

where ΔESS is the increase in the error sum of squares, k is the number of coefficients, and T is the number of observations. This contrasts with the classical summary statistic $\Delta ESS/\sigma^2$, which is comparedwith $\chi_p^2(\alpha)$ where p is the rank of $\mathbf{R}$ and α is the significance level. Thus the classical counterpart of (6.16) is the ratio $\Delta ESS/\sigma^2\chi_p^2(\alpha)$. In addition to the nonoccurrence of the factor T^{-1} (which for large T necessitates a "significant" finding), the classical summary differs from the subjectivist summary in depending on p, the number of restrictions. The measure (6.16), incidentally, is just the difference in the multiple correlation coefficients of the two models times a factor that tends to a constant as sample size grows, $L^2(\mathbf{R},\mathbf{r})=(R^2-R_0^2)(\mathbf{Y}'\mathbf{M}\mathbf{Y}/T\sigma^2)/(1+kT^{-1})$. Thus if a restriction does not greatly affect the R^2 of an equation, it will not greatly increase the expected squared prediction error.

This rough coincidence of approaches usefully highlights the assumptions that are implicit in the use of classical tests to simplify models for