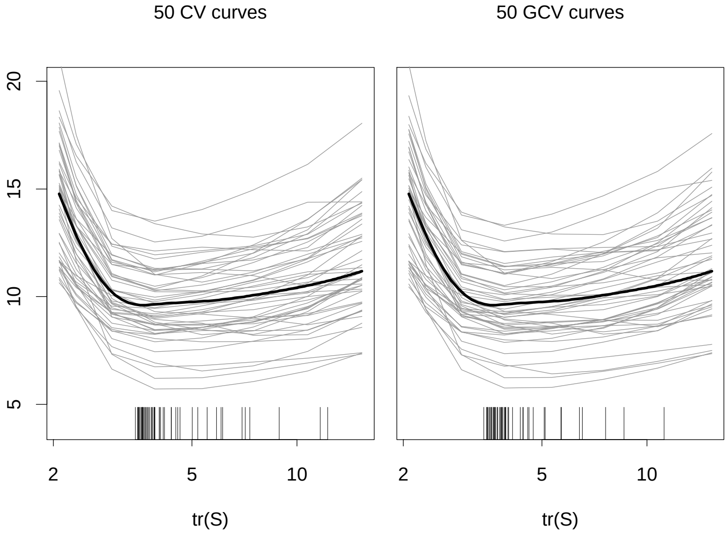

Fig. 3.3. *CV and GCV curves for 50 simulations from the same quadratic model as in Fig. 3.2. Here the smoother is a cubic smoothing spline. Each light curve represents a CV/GCV curve for one of the simulations. Each vertical line at the base of the plot corresponds to a minimum of a curve. We have plotted the curves as a function of* $\text{tr}(\mathbf{S}_\lambda)$ *(and on the log scale) rather than* λ *itself, since this is a more meaningful calibration. The solid curve is the true PSE curve.*

Our experience and that of others has indicated that GCV tends to undersmooth, and in these situations the GCV curve typically has multiple minima. Undersmoothing is particularly prevalent in small datasets, where short trends in the plot of Y against X are interpreted as high-frequency structure.

We have seen that generalized cross-validation can be viewed as an approximation to cross-validation. Here we show a simple way to compare C_p to GCV. Using the approximation $(1-x)^{-2} \approx 1+2x$ we obtain

$$GCV(\lambda) \approx \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}_\lambda(x_i)\}^2 + 2\text{tr}(\mathbf{S}_\lambda)\frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}_\lambda(x_i)\}^2.$$

Note that the right hand side is the same as the C_p statistic, except that the estimate $\frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}_\lambda(x_i)\}^2$ is used for $\hat{\sigma}^2$, while C_p uses a separate estimate based on a low bias smoother.