Note 1:

$$\hat{y} = X(X^TX + \lambda D)^{-1} X^T y$$

$X = X_T$ for truncated power functions

$$\hat{y} = X_T(X_T^T X_T + \lambda D)^{-1} X_T^T y$$

$$= X_T L_p L_p^{-1} (X_T^T X_T + \lambda D)^{-1} (L_p^T)^{-1} (L_p^T) X_T^T y$$

$$= X_B L_p^{-1} (X_T^T X_T + \lambda D)^{-1} (L_p^T)^{-1} X_B^T y$$

$$= X_B \left[L_p^T (X_T^T X_T + \lambda D) L_p \right]^{-1} X_B^T y$$

$$= X_B \left[L_p^T X_T^T X_T L_p + \lambda L_p^T D L_p \right]^{-1} X_B^T y$$

$$= X_B (X_B^T X_B + \lambda L_p^T D L_p)^{-1} X_B^T y$$

Note 2:

$$MSSE(\hat{f}) = E \sum_{i=1}^{n} \left[\hat{f}(x_i) - f(x_i) \right]^2$$

$$= \sum_{i=1}^{n} E\left[\hat{f}(x_i) - f(x_i) \right]^2 \text{; using the MSE decomposition}$$

$$= \sum_{i=1}^{n} \left[E\hat{f}(x_i) - f(x_i) \right]^2 + Var\left[\hat{f}(x_i) \right] \text{; using } \hat{f} = Ly = S_\lambda y$$

$$= \sum_{i=1}^{n} \left[E(Ly)_i - f_i \right]^2 + Var\left[(Ly)_i \right]$$

$$= \sum_{i=1}^{n} \left[E(Ly)_i - f_i \right]^2 + \left[Cov(Ly) \right]_{ii}$$

$$= \|(L-I)f\|^2 + tr[cov(Ly)]$$

$$= \|(L-I)f\|^2 + tr[L\, cov(y)\, L^T]; \text{ using } cov(y) = \sigma_\xi^2 I$$

$$= \|(L-I)f\|^2 + \sigma_\xi^2\, tr(LL^T)$$

Note 3:

$$y = f + \xi$$

$$\hat{y} = S_\lambda y = Ly$$

$$E[RSS] = E\|y - \hat{y}\|^2 = E\|(L-I)y\|^2$$

$$= E[y^T (L-I)^T (L-I) y]$$

$$= f^T (L-I)^T (L-I) f + \sigma_\xi^2\, tr[(L-I)^T (L-I)]$$

$$= \|(L-I)f\|^2 + \sigma_\xi^2 \left[tr[LL^T] - 2\, tr[L] + n \right]$$

$$= \|(L-I)f\|^2 + \sigma_\xi^2\, tr[LL^T] - 2\sigma_\xi^2\, tr(L) + n\sigma_\xi^2$$

$$= MSSE(\hat{f}) + \sigma_\xi^2 (n - 2\, df_{Fit}); \text{ where } df_{Fit} = tr(L).$$

Note 4

$$E[RSS] = \|(L-I)f\|^2 + \sigma_\xi^2\left[tr(LL^T) - 2tr(L) + n\right]$$

$$= \|(L-I)f\|^2 + \sigma_\xi^2 df_{res}$$

$$n - df_{res} = 2tr(L) - tr(LL^T)$$

Assuming the bias is negligible, $\|(L-I)f\|^2 \approx 0$

$$RSS/df_{res}$$

is an unbiased estimate of σ_ξ^2