1. Solving the square:

\[
\frac{1}{n} \sum_{i=1}^{n} (w\hat{e}_i + (1 - w) \tilde{e}_i)^2 = w^2 \frac{1}{n} \sum_{i=1}^{n} \tilde{e}_i^2 + 2w (1 - w) \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \tilde{e}_i + (1 - w)^2 \frac{1}{n} \sum_{i=1}^{n} \tilde{e}_i^2
\]

Now using matrix notation

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \tilde{e}_i = \hat{e}' \tilde{e} = \frac{1}{n} y' MM_1 y = \frac{1}{n} y' M y = \frac{1}{n} \hat{e}' \hat{e} = \hat{\sigma}^2
\]

where \( M = I - X (X'X)^{-1} X' \), \( M_1 = I - X_1 (X'_1X_1)^{-1} X'_1 \) and \( X = [X_1, X_2] \). The third inequality uses the fact that since \( X_1 \) lies in the span of \( X \), \( MM_1 = M \).

\[
y' (I - P - P_1 + P_1) y = y' (I - P) y = .
\]

This means that the top equation simplifies to

\[
(w^2 + 2w (1 - w)) \hat{\sigma}^2 + (1 - w)^2 \tilde{\sigma}^2 = (1 - a) \hat{\sigma}^2 + a \tilde{\sigma}^2
\]

if we set

\[
a = (1 - w)^2,
\]

noting that

\[
1 - (1 - w)^2 = w^2 + 2w (1 - w).
\]

2.

\[
\frac{1}{n} \text{tr} (MD) = \frac{1}{n} \text{tr} \left( (I - X (X'X)^{-1} X') D \right) = \frac{1}{n} \text{tr} (D) - \text{tr} \left( X (X'X)^{-1} X'D \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 - \frac{1}{n} \text{tr} \left( \left( \frac{1}{n} X'X \right)^{-1} \left( \frac{1}{n} X'DX \right) \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 - \frac{1}{n} b_n
\]

where

\[
b_n = \text{tr} \left( \left( \frac{1}{n} X'X \right)^{-1} \left( \frac{1}{n} X'DX \right) \right).
\]
Now

\[ \frac{1}{n} X'X = \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \xrightarrow{p} E(x_i x_i') = Q \]

\[ \frac{1}{n} X'DX = \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \sigma_i^2 \xrightarrow{p} E(x_i x_i' \sigma_i^2) \]

and

\[ E(x_i x_i' \sigma_i^2) = E(x_i x_i' E(e_i^2 | x_i)) = E(E(x_i x_i' e_i^2 | x_i)) = E(x_i x_i' e_i^2) = \Omega. \]

Thus \[ b_n \xrightarrow{p} \text{tr} (Q^{-1} \Omega). \]

Cautionary Remark: Common mistakes include mis-using the matrix manipulations, substituting between \( \sigma_i^2, e_i^2 \) and \( \hat{e}_i^2 \) in definitions.

3. A good test for \( H_0 \) is the Wald test. This is appropriate in this context because the estimator is least-squares and the hypothesis is a linear restriction on the least-squares coefficients. Let \( k = \dim(\beta_1) = \dim(\beta_2). \) (Note that \( \beta_1 \) and \( \beta_2 \) must have the same number of elements otherwise the hypothesis does not make sense. Furthermore, while the dimensions of \( \beta_1 \) and \( \beta_2 \) are not given, the use of the transpose operator clearly suggests that they are vectors, not scalars. Thus it is inappropriate to simply assume that \( k = 1, \) which was a common error.)

The hypothesis can be written in the linear form

\[ H_0 : R' \beta = 0 \]

where

\[ R = \begin{pmatrix} I_k & -I_k \end{pmatrix}. \]

First, estimate the model by least-squares. The coefficient estimate is

\[ \hat{\beta} = (X'X)^{-1} (X'y). \]

An estimate of the asymptotic covariance matrix is

\[ \hat{\Sigma} = \hat{Q}^{-1} \hat{\Omega} \hat{Q}^{-1} \]

\[ \hat{Q} = \frac{1}{n} X'X \]

\[ \hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \hat{e}_i^2 \]

\[ \hat{e}_i = y_i - x_i \hat{\beta} \]
The Wald statistic for $H_0$ against $H_1$ is

$$W_n = n \hat{\beta}' R (R' \hat{V} R)^{-1} R' \hat{\beta}$$

You can also write it as

$$W_n = n (\hat{\beta}_1 - \hat{\beta}_2)' \left( \hat{V}_{11} - \hat{V}_{21} - \hat{V}_{12} + V_{22} \right)^{-1} (\hat{\beta}_1 - \hat{\beta}_2).$$

I would select the 5% level, and would use an asymptotic test. The asymptotic test would reject $H_0$ in favor of $H_1$ if $W_n$ exceeds the 5% critical value of the $\chi^2_k$ distribution. The $\chi^2_k$ distribution is used since $W_n$ is asymptotically $\chi^2_k$ under $H_0$.

Alternatively, I could use a bootstrap test.

Since the model does not assume that the error is homoskedastic, it would be inappropriate to use an F statistic, or a Wald statistic constructed using the homoskedastic covariance matrix estimator.

4. Studying a list of t-ratios to find "the key predictor" can be quite misleading. One way to understand this deficiency is that the researcher is effectively looking for the largest t-ratio in a set of 20. While any individual t-ratio might be approximately normally distributed, the maximum over a set of 20 t-ratios is not normally distributed. It should not be surprising to see one "significant" t-ratio among a set of 20, even if all the coefficients are truly zero.

To make a formal argument, suppose all of the coefficients are zero and the covariance matrix $V$ is diagonal, so the t-ratios $T_j \to d Z_j \sim N(0, 1)$ and are mutually independent. We can then calculate the probability that one of the 20 t-ratios is "significant" in the sense of exceeding 2 in absolute value. We calculate that

$$P \left( \max_{1 \leq j \leq 20} |T_j| \leq 2 \right) \to P \left( \max_{1 \leq j \leq 20} |Z_j| \leq 2 \right) = P (|Z_j| \leq 2)^{20} = (0.95)^{20} = 0.36.$$  

Therefore, under the hypothesis that all coefficients are zero, the probability that the largest t-ratio exceeds 2 is approximately 64%. It is hardly surprising that this occurs. The researcher actually observed 2.5, which is somewhat less likely. Indeed

$$P \left( \max_{1 \leq j \leq 20} |T_j| \leq 2.5 \right) \to (0.99)^{20} = 0.82.$$  

Thus the probability of observing a t-ratio this large (under the hypothesis that all coefficients are truly zero) is approximately 18%. While somewhat unlikely, this is still insignificant by conventional standards.