How F Relates to t

Since the t-test is more intuitive than the F-test, it may help to see what the relationship is between the two tests. First, suppose we do the F-test with $p_0 = 1$. Then both tests are testing whether $\beta_p = 0$ (we are taking the last column, $X^{[p]}$, for convenience as the one we are testing), so they have the very same null hypothesis. Let W_{p-1} denote the column space of the first $p-1$ columns of X. Define $Z := X^{[p]} - P_{W_{p-1}} X^{[p]}$ to be the part of $X^{[p]}$ that is orthogonal to W_{p-1}. Then

$$\hat{Y} = X\hat{\beta} = \sum_{k=1}^{p-1} X^{[k]}\hat{\beta}_k + X^{[p]}\hat{\beta}_p = \sum_{k=1}^{p-1} X^{[k]}\hat{\beta}_k + \big(P_{W_{p-1}} X^{[p]} + Z\big)\hat{\beta}_p = \hat{Y}^{(\mathrm{s})} + Z\hat{\beta}_p \tag{1}$$

since all terms but the last belong to W_{p-1} and the last term is $\perp W_{p-1}$. Equation (1) shows that the numerator of F is $\|\hat{Y}\|^2 - \|\hat{Y}^{(\mathrm{s})}\|^2 = \|Z\hat{\beta}_p\|^2 = \|Z\|^2\hat{\beta}_p^2$, whence $F = \|Z\|^2\hat{\beta}_p^2/\hat{\sigma}^2$. Equation (1) also shows that $\hat{\beta}_p$ is the same for X as for the matrix V all of whose columns are the same as those of X except for the last one, which is changed to Z. You should check that since Z is orthogonal to the other columns of V, we have that all the entries of the last row and column of $V'V$ are 0 except for the (p,p)-entry, and that entry is $\|Z\|^2$. You should also check that this means that the (p,p)-entry of $(V'V)^{-1}$ is $1/\|Z\|^2$. Thus, we deduce that the SE of $\hat{\beta}_p$ is $\sigma^2/\|Z\|^2$. This shows that $t = \hat{\beta}_p/\big(\hat{\sigma}/\|Z\|\big) = \hat{\beta}_p\|Z\|/\hat{\sigma}$. Therefore, $F = t^2$.

That was for $p_0 = 1$. Now we derive a formula relating F to *several* t-statistics when $p_0 > 1$. Write $\hat{Y}^{(\mathrm{s},k)}$ for the fitted value of Y in the small model consisting of the first k columns of X. In this notation, $Y^{(\mathrm{s})} = Y^{(\mathrm{s},p_0)}$. Then we have a telescoping sum for the numerator of the numerator of F:

$$\begin{aligned}\|\hat{Y}\|^2 - \|\hat{Y}^{(\mathrm{s})}\|^2 &= \big(\|\hat{Y}\|^2 - \|\hat{Y}^{(\mathrm{s},p-1)}\|^2\big) + \big(\|\hat{Y}^{(\mathrm{s},p-1)}\|^2 - \|\hat{Y}^{(\mathrm{s},p-2)}\|^2\big) \\ &\quad + \big(\|\hat{Y}^{(\mathrm{s},p-2)}\|^2 - \|\hat{Y}^{(\mathrm{s},p-3)}\|^2\big) + \cdots + \big(\|\hat{Y}^{(\mathrm{s},p-p_0+1)}\|^2 - \|\hat{Y}^{(\mathrm{s})}\|^2\big)\,.\end{aligned}$$

Each of these terms can be treated as above where we had $p_0 = 1$. However, since we are using $\|e\|$ from the big model in the denominator of F, i.e., we are estimating σ from the big model (which makes sense since it gives us the most information about σ), the terms don't quite match those of the squares of the corresponding t-statistics, which are from various smaller models. But it makes sense to consider modified t-statistics, where we use the same $\hat{\sigma}$ always. Thus, let $t_{(\mathrm{s},k)} := \hat{\beta}_k^{(\mathrm{s},k)}/\widehat{\mathrm{SE}}_k$, where the numerator is the estimated coefficient of $X^{[k]}$ in the model from the first k columns of X and the denominator is the estimated SE of the numerator, using our fixed $\hat{\sigma} = \|e\|/\sqrt{n-p}$. This gives $F = (1/p_0)\sum_{k=p-p_0+1}^{p} t_{(\mathrm{s},k)}^2$. Thus, F is an *average* of modified t-statistics. It is possible to rederive the distribution of F from this formula.