Of course, these arguments prove nothing about admissibility but do suggest that the necessity for the known mean of the V_i's is not unreasonable.

SCHOOL OF MATHEMATICS AND STATISTICS
THE UNIVERSITY OF BIRMINGHAM
P.O. BOX 363
BIRMINGHAM B15 2TT
UNITED KINGDOM

B. EFRON

Stanford University

Here is a slightly simpler version of Brown's nice paradox: the statistician observes $X \sim N_p(\mu, I)$, $p \geq 3$, and also an integer J that equals $j = 1, 2, 3, \ldots, p$ with probability $1/p$, independently of X. It is desired to estimate μ_j with squared-error loss. Then J is ancillary, and conditional on $J = j$ the obvious estimate $d_0(X, j) = X_j$ is admissible and minimax. Unconditionally, however, the Jth coordinate of the James–Stein estimate,

$$d_1(X, J) = \left[1 - (p-2)/\|X\|^2\right] X_J,$$

dominates $d_0(X, J)$, with $E[d_1(X, J) - \mu_J]^2 < E[d_0(X, J) - \mu_J]^2$ for all vectors μ.

In other words, Brown has restated Stein's paradox, that d_1 dominates d_0 in terms of total squared error loss, in an interesting way that casts some doubt on the ancillarity principle.

[The example above does not look much like Brown's regression paradox, but we can fix things up by supposing that given $J = j$ the statistician also observes $X_0 \sim N(\alpha + \mu_j, 1)$, independent of $X \sim N_p(\mu, I)$, the goal now being to estimate α with squared-error loss. Then $\hat{\alpha}_1 = X_0 - d_1(X, J)$ dominates $\hat{\alpha}_0 = X_0 - d_0(X, J)$ unconditionally but not conditionally. This situation might arise if X_j was the placebo response of patient j on some physiological scale and X_0 was patient j's response when given a treatment of interest; we placebo-test p patients and then choose one at random to receive the treatment.]

Why do we intuitively accept the ancillarity principle in Cox's example, Section 5, but doubt it in the example above, or in Brown's regression paradoxes? I believe that the answer has more to do with single versus multiple inference than with hypothesis testing versus estimation.

Notice that $d_0(X, j)$ disregards all of the data except X_j. There is nothing in the ancillarity principle to justify this. All that ancillarity says is that we should do our probability calculations conditional on $J = j$. In Cox's example on the other hand, the conditional solution makes use of all the data and the ancillarity principle works fine.

Even when the choice $J = j$ is totally nonrandom it is not obvious that d_0 is preferable to d_1. The real question is whether or not the ensemble estimation gains offered by d_1 are relevant to the specific problem of estimating μ_j.