The two terms of (11.20) are of different orders. The first term is a weighted sum of the elements of the vector $\boldsymbol{u}$, each of which has mean zero. Thus, under suitable regularity conditions, it is easy to see that

$$n^{-1/2}\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{u} \overset{a}{\sim} N\Big(\boldsymbol{0},\ \operatorname*{plim}_{n\to\infty}\big(n^{-1}\sigma_1^2\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{X}\boldsymbol{\beta}\big)\Big).$$

This first term is thus $O(n^{1/2})$. The second term, in contrast, is $O(1)$, since

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\big(\boldsymbol{u}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{u}\big) &= \operatorname*{plim}_{n\to\infty}\big(\boldsymbol{u}^{\top}\boldsymbol{P}_Z\boldsymbol{u} - \boldsymbol{u}^{\top}\boldsymbol{P}_Z\boldsymbol{P}_X\boldsymbol{u}\big)\\ &= \sigma_1^2 k_2 - \sigma_1^2\lim_{n\to\infty}\big(\mathrm{Tr}\,(\boldsymbol{P}_Z\boldsymbol{P}_X)\big),\end{aligned}$$

and the trace of $\boldsymbol{P}_Z\boldsymbol{P}_X$ is $O(1)$. Thus, asymptotically, it is only the first term in (11.20) that matters.

Similarly, under H_1 the factor in parentheses in the denominator of (11.18) is equal to

$$\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{u} + \boldsymbol{u}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{u}. \tag{11.21}$$

By arguments similar to those used in connection with the numerator, the first of the three terms in (11.21) may be shown to be $O(n)$, the second $O(n^{1/2})$, and the third $O(1)$. Moreover, it is clear that $\acute{s} \to \sigma_1$ under H_1. Thus, asymptotically under H_1, the test statistic (11.18) tends to the random variable

$$\frac{\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{u}}{\sigma_1\big(\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{X}\boldsymbol{\beta}\big)^{1/2}},$$

which can be shown to be distributed asymptotically as $N(0,1)$.

This analysis not only makes it clear why the J and P tests are valid asymptotically but also indicates why they may not be well behaved in finite samples. When the sample size is small or $\boldsymbol{Z}$ contains many regressors that are not in $\mathcal{S}(\boldsymbol{X})$, the quantity $\boldsymbol{u}^{\top}\boldsymbol{P}_Z\boldsymbol{M}_X\boldsymbol{u}$, which is asymptotically negligible, may actually be large and positive. Hence, in such circumstances, the J-test statistic (11.18) may have a mean that is substantially greater than zero.

Several ways of reducing or eliminating this bias have been suggested. The simplest, which was first proposed by Fisher and McAleer (1981) and further studied by Godfrey (1983), is to replace $\hat{\boldsymbol{\gamma}}$ in the J-test and P-test regressions by $\tilde{\boldsymbol{\gamma}}$, which is the estimate of $\boldsymbol{\gamma}$ obtained by minimizing

$$\big(\hat{\boldsymbol{x}} - \boldsymbol{z}(\boldsymbol{\gamma})\big)^{\top}\big(\hat{\boldsymbol{x}} - \boldsymbol{z}(\boldsymbol{\gamma})\big).$$

Thus $\tilde{\boldsymbol{\gamma}}$ is the NLS estimate of $\boldsymbol{\gamma}$ obtained when one uses the fitted values $\hat{\boldsymbol{x}}$ instead of the dependent variable $\boldsymbol{y}$. In the linear case, this means that the J-test regression (11.16) is replaced by the regression

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \alpha\boldsymbol{P}_Z\boldsymbol{P}_X\boldsymbol{y} + \text{residuals}. \tag{11.22}$$

This regression yields what is called the **J_A test** because Fisher and McAleer attributed the basic idea to Atkinson (1970). Godfrey (1983) showed, using a result of Milliken and Graybill (1970), that the t statistic on the estimate of α from regression (11.22) actually has the t distribution in finite samples under the usual conditions for t statistics to have this distribution ($\boldsymbol{u}$ normally distributed, $\boldsymbol{X}$ and $\boldsymbol{Z}$ independent of $\boldsymbol{u}$). The intuition for this result is quite simple. The vector of fitted values $\boldsymbol{P}_X\boldsymbol{y}$ contains only the part of $\boldsymbol{y}$ that lies in $\mathcal{S}(\boldsymbol{X})$. It must therefore be independent of $\boldsymbol{M}_X\boldsymbol{y}$, which is what the residuals from (11.22) would be if $\alpha = 0$. Therefore, we can treat $\boldsymbol{P}_Z\boldsymbol{P}_X\boldsymbol{y}$ (or any other regressor that depends on $\boldsymbol{y}$ only through $\boldsymbol{P}_X\boldsymbol{y}$) as if it were a fixed regressor.[4] The **P_A test** is to the P test as the J_A test is to the J test.

Unfortunately, the J_A and P_A tests are in many circumstances much less powerful than the ordinary J and P tests; see Davidson and MacKinnon (1982) and Godfrey and Pesaran (1983). Thus if, for example, the J test rejects the null hypothesis and the J_A test does not, it is hard to know whether this is because the former is excessively prone to commit a Type I error or because the latter is excessively prone to commit a Type II error.

A second approach is to estimate the expectation of $\boldsymbol{u}^\top\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{u}$, subtract it from $\boldsymbol{y}^\top\boldsymbol{M}_X\boldsymbol{P}_Z\boldsymbol{y}$, and then divide it by an estimate of the square root of the variance of the resulting quantity so as to obtain a test statistic that would be asymptotically $N(0, 1)$. This approach was originally proposed in a somewhat more complicated form by Godfrey and Pesaran (1983); a simpler version may be found in the "Reply" of MacKinnon (1983). This second approach is a good deal harder to use than the J_A test, since it involves matrix calculations that cannot be performed by a sequence of regressions, and it does not yield an exact test. It also requires the assumption of normality. However, it does seem to yield a test with much better finite-sample properties under the null than the J test and, at least in some circumstances, much better power than the J_A test.

The vector $\tilde{\boldsymbol{\gamma}}$ is of interest in its own right. The original Cox test used the fact that, under H_1,

$$\operatorname*{plim}_{n\to\infty}\bigl(\tilde{\boldsymbol{\gamma}}\bigr) = \operatorname*{plim}_{n\to\infty}\bigl(\hat{\boldsymbol{\gamma}}\bigr).$$

It is possible to construct a test based directly on the difference between $\hat{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}}$. Such a test, originally proposed by Dastoor (1983) and developed further by Mizon and Richard (1986), looks at whether the value of $\boldsymbol{\gamma}$ predicted by the H_1 model (i.e., $\tilde{\boldsymbol{\gamma}}$) is the same as the value obtained by direct estimation of H_2 (i.e., $\hat{\boldsymbol{\gamma}}$). These tests are called **encompassing tests**, because if H_1 does explain the performance of H_2, it may be said to "encompass" it; see Mizon (1984). The principle on which they are based is sometimes called the **encompassing principle**.

[4] By the same argument, the RESET test discussed in Section 6.5 is exact in finite samples whenever an ordinary t test would be exact.