Another Look at Linear Hypothesis Testing in Dense High-Dimensional Linear Models

Ronald Christensen*

August 13, 2018

Abstract

Zhu and Bradic (2017) presented methods for testing a one-dimensional linear hypothesis in linear models that have $p >> n$. Their statistics take the form of t statistics with asymptotic normal distributions. To deal with $p >> n$, strong additional assumptions have to be made beyond those made for traditional linear models. One assumption they make is that of a random design.

I present alternative formulations of their problems that suggest some attractive alternatives to the numerators in their statistics and some clearly superior denominators. I also present some approaches for nonrandom designs while drawing attention to the strong additional assumptions required to solve the problem.

*Ronald Christensen is a Professor in the Department of Mathematics and of Statistics, University of New Mexico, Albuquerque, NM, 87131.

1 Introduction

Zhu and Bradic (2017) (henceforward ZB) consider the problem of fitting a linear model

$$y_i = x_i'\beta_* + \varepsilon_i, \quad \mathrm{E}(\varepsilon_i) = 0, \quad \mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2, \quad i = 1, \ldots, n$$

with (presumably) iid errors and testing an hypothesis

$$a'\beta_* = g_0.$$

Here the x_i and a vectors are p dimensional and ZB are primarily concerned with $p >> n$. Rewrite the linear model in matrix form as

$$Y = X\beta_* + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma_\varepsilon^2 I$$

with

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}, \quad e = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The solution to the testing problem is well-known in linear model theory, cf. Christensen (2011, Chapter 3). If $a'\beta_*$ is estimable, the result is straight-forward and if $a'\beta_*$ is not estimable, *without making additional assumptions* the problem is nonsensical. This is true regardless of the sizes of p and n. When $p >> n$, the p-dimensional vectors a defining an estimable function are restricted to a subspace of $\mathbf{R}^p$ with dimension $rank(X) \leq n << p$, so the estimability restriction is considerable. I will return to this issue in Sections 3 and 4. While ZB do an extensive literature review, they do not cite any books on linear model theory.

ZB consider a random design setting with $\mathrm{E}(x_i) = 0$ and $\mathrm{Cov}(x_i) = \Sigma_X$. They implicitly assume that the ε_is are independent of the x_is and that Σ_X is positive definite. (They write $\Sigma_X \equiv \Omega^{-1}$ but I will not use the Ω notation.) Note that in this random design model, unconditionally

we must have $\mathrm{E}(y_i) = 0$. It is not hard to deal with the more general cases having $\mathrm{E}(x_i) = \mu_X$ and $\mathrm{E}(y_i) = \mu_Y$, you merely include an intercept in the model and then do all the computations after subtracting the sample means. ZB's tests are not conditional on the x_is, as typical linear model tests are.

Because of the estimability issue, it has been clear to me for about 25 years, when I was first introduced to multivariate calibration problems (e.g., Fugate et al. 2002), that to deal effectively with $p >> n$ one needs more information than is available in a standard linear model. In particular, the most fruitful path seems to be having some kind of model for Σ_X. In their Section 2, ZB assume that Σ_X is known and, in my Section 2, I propose alternatives to their t-like statistics and discuss the merits of the alternative tests. These include more attractive/simpler numerators and clearly better denominators.

In Section 3 we will see that ZB's approach to Σ_X unknown involves extending the ideas of Section 2 by incorporating an alternative strong assumption. Again, I present simpler numerators and clearly better denominators.

Section 4 presents methods that do not require the random design assumption but that involve alternative strong additional assumptions.

2 Testing $H_0 : a'\beta_* = g_0$ with prior knowledge of Σ_X

In this section I propose several alternative test statistics two of which seem to be clear improvements on the statistic $T_n(g_0)$ proposed by ZB. One, t_1, is a more direct application of basic statistical principals. Another, t_5, is a relatively minor modification of $T_n(g_0)$ that clearly provides more power.

2.1 Proposed Tests

I will derive test statistics using the theories of linear models and best linear prediction, cf. Christensen (2011). The proposed tests are straight-forward applications of the standard ideas of iden-

tifying a parameter Par, an estimate of that parameter Est, finding an observable standard error for Est, say $\mathrm{SE}(ESt)$, and using either normality or asymptotics to find the reference distribution $(Est - Par)/\mathrm{SE}(Est) \sim t(df)$. When appealing to asymptotics $t(\infty) \sim N(0,1)$. The test of $Par = m_0$, simply replaces Par with m_0, making for an observable test statistic that can be compared to the reference distribution.

In ZB's random design setting the linear model is closely related to best linear prediction. In particular, the vector of regression coefficients is

$$\beta_* = \Sigma_X^{-1}\Sigma_{XY},$$

where $\Sigma_{XY} \equiv \mathrm{Cov}(x_i, y_i)$ is unknown. The hypothesis now reduces to

$$a'\Sigma_X^{-1}\Sigma_{XY} = g_0$$

where everything but Σ_{XY} is known. An obvious estimate of Σ_{XY}, when the x_is have mean zero, is $S_{XY} \equiv \frac{1}{n}X'Y$.

The parameter is $Par = a'\Sigma_X^{-1}\Sigma_{XY}$. The hypothesized value for the parameter is g_0. The estimate of the parameter is $Est = a'\Sigma_X^{-1}X'Y/n = a'\Sigma_X^{-1}S_{XY}$. This estimate is unconditionally unbiased but not conditionally unbiased. Tarpey et al. (2014, 2015), Cook, Forzani, and Rothman (2013, 2014), and Ding (2014, 2015) all discuss inadequacies of this estimate, as opposed to the least squares estimate, but the least squares estimate only addresses the testing problem when $a'\beta_*$ is estimable, which is a big restriction when $p >> n$. As always, the most difficult part of developing a test is finding an appropriate standard error.

The simplest way to proceed with an unconditional test is to recognize that, with $\tilde{w}_i \equiv a'\Sigma_X^{-1}x_iy_i$, the proposed estimate is the sample mean of the $\tilde{w}_i$s which are iid. The unconditional standard error can be obtained from the sample variance of the $\tilde{w}_i$s as $\sqrt{s^2_{\tilde{w}}/n}$. For what it is worth,

$$s^2_{\tilde{w}} = \frac{1}{n}\left[a'\Sigma_X^{-1}X'D^2(Y)X\Sigma_X^{-1}a\right] - \left(a'\Sigma_X^{-1}S_{XY}\right)^2,$$

where $D(v)$ denotes a diagonal matrix with entries determined by the vector v. (Giving up the mean zero assumptions would invalidate the iid assertion.) This leads to the intuitively appealing test statistic

$$t_1 \equiv \frac{a'\Sigma_X^{-1}S_{XY} - g_0}{\sqrt{s_{\hat{w}}^2/n}}.$$

If the predictor variables are multivariate normal, a different unconditional test is possible because, cf. Tarpey et al. (2015),

$$\mathrm{Var}(a'\Sigma_X^{-1}S_{XY}) = a'\frac{\sigma_y^2\Sigma_X^{-1} + \Sigma_X^{-1}\Sigma_{XY}\Sigma_{XY}'\Sigma_X^{-1}}{n}a.$$

The unknown values σ_y^2 and Σ_{XY} are easily estimated. Multivariate normality is crucial because the variance is computed using the covariance matrix of a Wishart distribution. (The denominator of the variance would be $n-1$ without all the zero mean assumptions.) This leads to the test statistic

$$t_2 \equiv \frac{a'\Sigma_X^{-1}S_{XY} - g_0}{\sqrt{\left(s_y^2 a'\Sigma_X^{-1}a + a'\Sigma_X^{-1}S_{XY}S_{XY}'\Sigma_X^{-1}a\right)/n}}.$$

For $p > n$, the corresponding formula for the unconditional variance of the least squares estimate involves a denominator that is negative, so obviously the assumptions behind the computation have broken down and the least squares estimate cannot be used, cf. Tarpey et al. (2015).

Typically we prefer to analyze regression models conditionally. Conditional variances are never larger than unconditional ones. To get a conditional standard error, observe that conditional on X,

$$\mathrm{Var}(a'\Sigma_X^{-1}X'Y/n) = \frac{1}{n^2}a'\Sigma_X^{-1}X'(\sigma_\varepsilon^2 I)X\Sigma_X^{-1}a = \sigma_\varepsilon^2\frac{1}{n^2}a'\Sigma_X^{-1}X'X\Sigma_X^{-1}a = \sigma_\varepsilon^2\frac{1}{n}a'\Sigma_X^{-1}S_X\Sigma_X^{-1}a$$

where $S_X \equiv (1/n)X'X$. When $p >> n$, since S_X is a (not particularly good) estimate of Σ_X, e.g. S_X is not invertible, there *may* be advantages to replacing the estimate S_X with the known

parameter Σ_X and using

$$\mathrm{Var}(a'\Sigma_X^{-1}X'Y/n) \doteq \sigma_\varepsilon^2 \frac{1}{n} a'\Sigma_X^{-1}a,$$

which is the the expected value of the conditional variance of the estimate (but is not the unconditional variance).

The standard error will be an estimated standard deviation so we also need to estimate σ_ε^2. Fortunately, with a random design, σ_ε^2 is just the best linear prediction error variance

$$\sigma_\varepsilon^2 = \sigma_y^2 - \Sigma'_{XY}\Sigma_X^{-1}\Sigma_{XY}$$

and the only term that is difficult to estimate is known. Thus, with the xs and ys having mean 0,

$$s_\varepsilon^2 \equiv \frac{1}{n}\left[Y'Y - Y'X\Sigma_X^{-1}X'Y/n\right] = s_y^2 - S'_{XY}\Sigma_X^{-1}S_{XY}.$$

The usual test statistic is the estimate minus the hypothesized value divided by the standard error

$$t_3 \equiv \frac{a'\Sigma_X^{-1}S_{XY} - g_0}{\sqrt{s_\varepsilon^2 \frac{1}{n} a'\Sigma_X^{-1}S_X\Sigma_X^{-1}a}}$$

or alternatively, using the approximate variance,

$$t_4 \equiv \frac{a'\Sigma_X^{-1}S_{XY} - g_0}{\sqrt{s_\varepsilon^2 \frac{1}{n} a'\Sigma_X^{-1}a}}.$$

But again, these conditional test statistics are based on conditionally biased estimates of the parameter $a'\beta_*$.

2.2 ZB's Approach to Testing

We now consider two more test statistics, one of which has the form of the ZB test.

Consider a random sample $w_1, \ldots, w_n$ with $\mathrm{E}(w_i) = \mu$ and $\mathrm{Var}(w_i) = \sigma^2$. Let J denote an $n \times 1$

vector of 1s and write the corresponding linear model

$$W = J\mu + e.$$

The standard t statistic for testing $\mu = 0$ is

$$t_5 \equiv \frac{\bar{w}_{\cdot}}{\sqrt{s_w^2/n}} = \frac{\sqrt{n(n-1)}}{n} \frac{J'W}{\sqrt{W'[I - (1/n)JJ']W}},$$

which uses the variance estimate from the original (full) model.

Alternatively, one could estimate the variance using the (reduced) model that incorporates the null hypothesis $\mu = 0$. If $\mu = 0$, $\mathrm{E}(W'W) = n\sigma^2$ so define $\hat{\sigma}_w^2$ and the analogous test statistic

$$t_6 \equiv \frac{\bar{w}_{\cdot}}{\sqrt{\hat{\sigma}_w^2/n}} = \frac{J'W}{\sqrt{W'W}}.$$

As seen in the next subsection, it is t_6 that ZB employ.

In the one-sample problem, nobody uses t_6 and it is pretty easy to see that t_6 is not going to work as well as t_5. Just look at what happens when you replace the linear and quadratic statistics with their expectations. The basic computations are $\mathrm{E}(J'W) = n\mu$, $\mathrm{Var}(J'W) = n\sigma^2$, $\mathrm{E}\{W'[I - (1/n)JJ']W\} = (n-1)\sigma^2$, $\mathrm{E}(W'W) = n\sigma^2 + n\mu^2$. These lead to

$$|t_5| \doteq \frac{\sqrt{n(n-1)}}{n} \frac{n|\mu|}{\sqrt{(n-1)\sigma^2}} = \sqrt{n}\frac{|\mu|}{\sigma}$$

and

$$|t_6| \doteq \frac{n|\mu|}{\sqrt{n\sigma^2 + n\mu^2}} = \sqrt{n}\frac{|\mu|}{\sqrt{\sigma^2 + \mu^2}}.$$

As desired, when $\mu \neq 0$, both values go to infinity as $n \to \infty$. But the $|t_5|$ approximation is always larger than the $|t_6|$ approximation. For example, take $\mu = 3$ and $\sigma^2 = 1$, then $t_5 \doteq \sqrt{n}3$ but $t_6 \doteq \sqrt{n}\frac{3}{\sqrt{10}} < \sqrt{n}$.

2.3 Comparison of Tests

ZB's test statistic is

$$T_n(g_0) = \frac{\sum_i z_i(y_i - z_i g_0)}{\sqrt{\sum_i [z_i(y_i - z_i g_0)]^2}}$$

where $(z_1, \ldots, z_n)' \equiv Z$ is defined as

$$Z \equiv X\Sigma_X^{-1} a (a'\Sigma_X^{-1} a)^{-1}$$

and in particular

$$z_i(y_i - z_i g_0) = (a'\Sigma_X^{-1} a)^{-1} a'\Sigma_X^{-1} x_i y_i - (a'\Sigma_X^{-1} a)^{-2} a'\Sigma_X^{-1} x_i x_i'\Sigma_X^{-1} a g_0.$$

If we define $W \equiv (w_1, \ldots, w_n)'$ with

$$w_i \equiv z_i(y_i - z_i g_0),$$

then, as in the previous subsection, the ZB test statistic is

$$T_n(g_0) = \frac{J'W}{\sqrt{W'W}} = t_6. \tag{2.1}$$

Since the (y_i, x_i')s are iid under the random design assumption, so are the w_is, so $\mathrm{Cov}(W) = \sigma^2 I$ for some σ^2. To see that $\mathrm{E}(W) = 0$ under the null model,

$$\begin{aligned}
\mathrm{E}(w_i) &= \mathrm{E}\left[z_i(y_i - z_i g_0)\right] \\
&= (a'\Sigma_X^{-1} a)^{-1} \mathrm{E}\left[a'\Sigma_X^{-1} x_i y_i - (a'\Sigma_X^{-1} a)^{-1} a'\Sigma_X^{-1} x_i x_i'\Sigma_X^{-1} a g_0\right] \\
&= (a'\Sigma_X^{-1} a)^{-1} \left[a'\Sigma_X^{-1}\Sigma_{XY} - (a'\Sigma_X^{-1} a)^{-1} a'\Sigma_X^{-1}\Sigma_X\Sigma_X^{-1} a g_0\right] \\
&= (a'\Sigma_X^{-1} a)^{-1} \left[a'\beta_* - (a'\Sigma_X^{-1} a)^{-1} a'\Sigma_X^{-1} a g_0\right] = 0.
\end{aligned}$$

Of course this is the unconditional expectation that depends crucially on the random design

assumption.

If the hypothesis is false, say $a'\beta_* = g_0 + d$, then $\mathrm{E}(w_i) = (a'\Sigma_X^{-1}a)^{-1}d$, so as in the previous subsection a one-sample linear model applies. An obvious improvement for $T_n(g_0)$ is to replace its denominator and use t_5.

We now compare the numerator of $T_n(g_0)$ to that of the t_1 through t_4 statistics. The numerator is

$$\begin{aligned} Z'(Y - Zg_0) &= Z'Y - Z'Zg_0 \\ &= (a'\Sigma_X^{-1}a)^{-1}a'\Sigma_X^{-1}X'Y - (a'\Sigma_X^{-1}a)^{-1}a'\Sigma_X^{-1}X'X\Sigma_X^{-1}a(a'\Sigma_X^{-1}a)^{-1}g_0 \\ &= (a'\Sigma_X^{-1}a)^{-1}\left\{a'\Sigma_X^{-1}X'Y - \left[(a'\Sigma_X^{-1}a)^{-1}a'\Sigma_X^{-1}X'X\Sigma_X^{-1}a\right]g_0\right\} \\ &= n(a'\Sigma_X^{-1}a)^{-1}\left(\left\{a'\Sigma_X^{-1}S_{XY} - \left[(a'\Sigma_X^{-1}a)^{-1}a'\Sigma_X^{-1}S_X\Sigma_X^{-1}a\right]g_0\right\}\right). \end{aligned}$$

If $S_X \to \Sigma_X$, then $\left[(a'\Sigma_X^{-1}a)^{-1}a'\Sigma_X^{-1}S_X\Sigma_X^{-1}a\right] \to 1$, so this would converge to $n(a'\Sigma_X^{-1}a)^{-1}$ times the numerator of the statistics in Subsection 2.1 (that have a more direct appeal).

I mentioned earlier the need to model Σ_X. In applications to spatial, temporal, multivariate, and longitudinal data, appropriate generalized least squares models, say,

$$Y = X\beta_* + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma_\varepsilon^2 V,$$

are complicated by the need to model V as some function of a parameter vector (e.g. Christensen, 2001). Similarly, the testing results examined here should get more complicated if one chooses a parametric model for Σ_X. It seems unrealistic to hope that the problem of testing with $p >> n$ has a good solution without restricting the number of unknown parameters to be substantially less than the number of observations (regardless of whether you consider the number of observations n or $(p+1) \times n$). When that is not true, the only realistic hope for a good solution is a Bayesian solution with very good prior information.

2.4 Transformation Motivation

We now provide some motivation for the ZB testing procedure that will inform our discussion of procedures for the case with Σ_X unknown. The idea is to transform the original linear model into a one-sample problem that involves a mean related to the parameter $a'\beta_*$.

The original linear model is

$$Y = X\beta_* + e.$$

Let A be an oblique projection operator onto $C(a)$, the column space of a, defined by $A \equiv a(a'\Sigma_X^{-1}a)^{-1}a'\Sigma_X^{-1}$. Note that $Za' = XA'$. Clearly, we can write

$$\begin{aligned} Y &= XA'\beta_* + X(I-A)'\beta_* + e \\ &= Z(a'\beta_*) + X(I-A)'\beta_* + e. \end{aligned}$$

Moreover,

$$Y - Zg_0 = Z(a'\beta_* - g_0) + X(I-A)'\beta_* + e,$$

where the dependent variable vector $Y - Zg_0$ is observable.

In order to create a test for the parameter $a'\beta_* - g_0$ we want to multiply the model by some matrix Q such that $\mathrm{E}[QX(I-A)'] = 0$ and $\mathrm{E}(QZ) = kJ$ for some scalar $k \neq 0$ in order to get a one-sample model

$$Q(Y - Zg_0) = J(a'\beta_* - g_0)k + Qe$$

that provides an easy test of $H_0 : (a'\beta_* - g_0) = 0$. If $Q = D(q)$, where the elements of q are $q_i = f(y_i, x_i')$ for some function f, the elements of $Q(Y - Zg_0)$ will be iid.

Fortunately, as partially demonstrated earlier, taking $Q = D(Z)$ does the trick. In particular, with e independent of X and thus independent of Z,

$$\mathrm{E}[D(Z)(Y - Zg_0)] = \mathrm{E}[D(Z)Z(a'\beta_* - g_0)] + \mathrm{E}[D(Z)X(I-A)'\beta_*] + \mathrm{E}[D(Z)e]$$

$$
\begin{aligned}
&= \mathrm{E}[D(Z)Z(a'\beta_* - g_0)] + \frac{a'\Sigma_X^{-1}\Sigma_X(I-A)'}{a'\Sigma_X^{-1}a}\beta_* + 0 \\
&= \mathrm{E}[D(Z)Z(a'\beta_* - g_0)] + 0 \\
&= \frac{a'\beta_* - g_0}{a'\Sigma_X^{-1}a} J.
\end{aligned}
$$

3 Testing $H_0 : a'\beta_* = g_0$ without prior knowledge of Σ_X

Testing without knowledge of Σ_X requires modification of the transformation method just discussed and requires additional assumptions to justify fitting linear models using penalized least squares estimates.

The linear model is

$$Y = X\beta_* + e.$$

Let M_a be the perpendicular projection operator onto $C(a)$ defined by $M_a \equiv a(a'a)^{-1}a'$. Redefine Z so that $Za' \equiv XM_a$. Also take U_a to be a matrix with orthonormal columns and $U_aU_a' = (I - M_a)$. In particular, $C(U_a) = C(a)^\perp = C(I - M_a)$. Unrelated to W and $\tilde{w}_i$ in the previous section define $\tilde{W} \equiv XU_a$. Both $\tilde{W}$ and $X(I - M_a)$ are model matrices for a reduced model associated with tests involving $a'\beta_*$, cf. Christensen (2011, Section 3.3). Of course when $a'\beta_*$ is not estimable, $C(\tilde{W}) = C(X)$ and the hypothesis $a'\beta_* = g_0$ places no restriction on the linear model, so some other assumptions need to come into play.

Clearly, we can write

$$
\begin{aligned}
Y &= XM_a\beta_* + X(I - M_a)\beta_* + e \\
&= Z(a'\beta_*) + \tilde{W}\pi_* + e,
\end{aligned}
$$

where $\pi_* \equiv U_a'\beta_*$. Moreover,

$$Y - Zg_0 = Z(a'\beta_* - g_0) + \tilde{W}\pi_* + e, \tag{3.1}$$

where $Y - Zg_0$ is an observable vector of dependent variables.

Once again, in order to create a test for the parameter $a'\beta_* - g_0$, we want to multiply the model by some diagonal matrix Q such that $\mathrm{E}[Q\tilde{W}] = 0$ and $\mathrm{E}(QZ) \in C(J)$. Without knowing Σ_X, this requires another strong assumption. ZB assume

$$Z = \tilde{W}\gamma + u, \quad \mathrm{E}(u) = 0, \quad \mathrm{Cov}(u) = \sigma_u^2 I, \tag{3.2}$$

where the u_is are iid and, crucially, u is independent of $\tilde{W}$ and the error e in the original linear model.

This time take $Q = D(Z - \tilde{W}\gamma) = D(u)$. In particular,

$$\begin{aligned}
&\mathrm{E}[D(Z - \tilde{W}\gamma)(Y - Zg_0)] \\
&= \mathrm{E}[D(Z - \tilde{W}\gamma)Z(a'\beta_* - g_0)] + \mathrm{E}[D(Z - \tilde{W}\gamma)\tilde{W}\pi_*] + \mathrm{E}[D(Z - \tilde{W}\gamma)e] \\
&= \mathrm{E}[D(u)Z(a'\beta_* - g_0)] + \mathrm{E}[D(u)\tilde{W}\pi_*] + \mathrm{E}[D(u)e] \\
&= \mathrm{E}[D(u)Z](a'\beta_* - g_0) + \mathrm{E}[D(u)]\mathrm{E}[\tilde{W}\pi_*] + \mathrm{E}[D(u)]\mathrm{E}[e] \\
&= \mathrm{E}[D(u)u](a'\beta_* - g_0) + 0 + 0 \\
&= J\sigma_u^2(a'\beta_* - g_0).
\end{aligned}$$

Again the rows of $D(Z - \tilde{W}\gamma)(Y - Zg_0)$ are iid so the methods of Subsection 2.2 can be used.

Of course the problem is that we do not know γ, so we need to replace it with a consistent estimate obtained from fitting model (3.2). For $p > n$, the least squares estimate is typically not consistent. In fact, when $a'\beta_*$ is not estimable, typically $C(\tilde{W}) = \mathbf{R}^n$, so the least squares estimate of $Z - \tilde{W}\gamma$ is always 0. (If, for example, the rows of X contain exact replicates, $C(\tilde{W}) \neq \mathbf{R}^n$.) In any case, the proposed test statistic is

$$t_5 = \frac{\sqrt{n(n-1)}}{n} \frac{J'W}{\sqrt{W'[I - (1/n)JJ']W}}, \quad \text{where } W = D(Z - \tilde{W}\hat{\gamma})(Y - Zg_0).$$

To get useful results one needs to make enough assumptions so that, say, some penalized least squares (regularized) estimate of γ will be consistent. Together with the linear model (3.2) for Z, these assumptions amount to modeling Σ_X. In particular, the use of penalized least squares involves putting a lot of faith in the penalty function, to the point where one might almost as well assume that the penalty function defines a prior distribution and develop a Bayesian analysis.

Rather than this method, ZB use a method based on

$$\begin{aligned}
&\mathrm{E}[D(Z-\tilde{W}\gamma)(Y-Zg_0-\tilde{W}\pi_*)] \\
&= \mathrm{E}[D(Z-\tilde{W}\gamma)Z(a'\beta_*-g_0)]+\mathrm{E}[D(Z-\tilde{W}\gamma)e] \\
&= J\sigma_u^2(a'\beta_*-g_0).
\end{aligned}$$

One could again treat this as a one-sample problem but ZB use a different variance estimate based on the assumption that $u_i=(z_i-\tilde{w}_i'\gamma_*)$ and $(y_i-z_ig_0-\tilde{w}_i\pi_*)$ are independent. The former has mean zero, as does the latter under the null hypothesis, so

$$\mathrm{Var}[(z_i-\tilde{w}_i'\gamma_*)(y_i-z_ig_0-\tilde{w}_i\pi_*)]=\mathrm{Var}(z_i-\tilde{w}_i'\gamma_*)\mathrm{Var}(y_i-z_ig_0-\tilde{w}_i\pi_*)=\sigma_u^2\sigma_\varepsilon^2.$$

Estimating each variance separately, σ_u^2 from (3.2) and σ_ε^2 from (3.1) with $a'\beta_*-g_0=0$, suggests a test statistic

$$\sqrt{n}\frac{J'D(Z-\tilde{W}\gamma)(Y-Zg_0-\tilde{W}\pi_*)}{\|Z-\tilde{W}\gamma\|\,\|Y-Zg_0-\tilde{W}\pi_*\|}$$

or, equivalently,

$$\sqrt{n}\frac{(Z-\tilde{W}\gamma)'(Y-Zg_0-\tilde{W}\pi_*)}{\|Z-\tilde{W}\gamma\|\,\|Y-Zg_0-\tilde{W}\pi_*\|}.$$

Unfortunately, the variance estimate for σ_ε^2 is only valid under the null model so it suffers from the same problems discussed in Section 2.

Another complication of ZB's method is that, in addition to estimating γ consistently, since π_* is actually unknown, π_* also requires consistent estimation. ZB estimate π_* from (3.1) with

$a'\beta_* - g_0 = 0$, i.e., the null model

$$Y - Zg_0 = \tilde{W}\pi_* + e.$$

This model has the same model matrix $\tilde{W}$ as (3.2), so it has the same problems in fitting it. ZB use regularization to estimate π_* and propose the test statistic

$$S_n \equiv \sqrt{n}\frac{(Z - \tilde{W}\hat{\gamma})'(Y - Zg_0 - \tilde{W}\hat{\pi}_2)}{\|Z - \tilde{W}\hat{\gamma}\| \, \|Y - Zg_0 - \tilde{W}\hat{\pi}_*\|}.$$

An obvious improvement is to estimate σ_ε with $\|Y - Zg_0 - Z\hat{\pi}_1 - \tilde{W}\hat{\pi}_2\|/\sqrt{n}$ where $\hat{\pi}_1$ and $\hat{\pi}_2$ are estimated from

$$Y - Zg_0 = Z\pi_1 + \tilde{W}\pi_2 + e$$

with a penalty function that only shrinks π_2. The fitted model does not involve the restricted parameters of the original linear model (3.1). It is at least as general as (3.1), so it should provide a reasonable estimate of σ_ε when (3.1) is true.

Using my simpler numerator and an improved denominator based on the independence assumptions made for model (3.2) leads to a test statistic for $H_0 : a'\beta_* - g_0 = 0$ of

$$\tilde{S} \equiv \sqrt{n}\frac{(Z - \tilde{W}\hat{\gamma})'(Y - Zg_0)}{\|Z - \tilde{W}\hat{\gamma}\| \, \|Y - Zg_0 - Z\hat{\pi}_1 - \tilde{W}\hat{\pi}_2\|}.$$

4 Nonrandom Predictors

I now present some alternative ideas for testing $a'\beta_* = g_0$ other than the random design transformation methods so effectively exploited by ZB. Again, these involve major additional assumptions some of which involve the appropriateness of penalized estimates. The most important assumption is that the coefficient of a predictor variable in a given linear model has the same meaning as the coefficient of that predictor variable in a larger linear model. This is an assumption that is frequently decried in standard linear model theory.

By way of explicating the proposed method based on regularized linear models, I begin with

a more traditional linear model approach that includes assumptions strong enough to solve the problem. We focus on the case where $a'\beta_*$ is not estimable because (a) when it is estimable a perfectly good theory already exists and (b) when $p >> n$ relatively few of the possible $a'\beta_*$ parameters are estimable.

I continue the notation of Section 3.

4.1 Reduced Linear Model Method

It is well known, e.g. Christensen (2011, Section 3.3), that a reduced model associated with testing $a'\beta_* = g_0$ is

$$(Y - Xb_*) = \tilde{W}\eta + e,$$

where b_* is any known (computed) solution $a'b_* = g_0$. In particular, $b_* \equiv a[g_0/(a'a)]$ provides a solution for which $Xb_* = Zg_0$. This is "the" reduced model in the sense that $C(\tilde{W})$ is uniquely determined. Again, the problem with $a'\beta_*$ not being estimable is that the model matrix column spaces have $C(\tilde{W}) = C(X)$, so the "reduced" model is actually equivalent to the original model.

We have rewritten the original linear model as

$$Y - Zg_0 = Z\delta + \tilde{W}\pi_* + e, \quad \delta \equiv (a'\beta_* - g_0), \quad \pi_* \equiv U_a'\beta_*. \tag{4.1}$$

This model has restrictions on the parameters δ and π_*, so it is apparently less general than an unrestricted model, say,

$$Y - Zg_0 = Z\pi_1 + \tilde{W}\pi_2 + e. \tag{4.2}$$

However, when $a'\beta_*$ is not estimable, $C(\tilde{W}) = C(X)$, so this unrestricted model is still equivalent to the original model. Regardless of estimability, there is no compelling reason to view π_1 as $a'\beta_* - g_0$ in the unrestricted model, but it is not a crazy thing to do and this section focuses on testing $\pi_1 = 0$.

Now lets put on an additional strong assumption of a reduced model

$$\mathrm{E}(Y - Zg_0) = Z\pi_1 + X_0\pi_3, \quad \pi_1 = (a'\beta_* - g_0), \quad C(X_0) \subset C(X), \quad C(X_0) \neq C(X). \tag{4.3}$$

Since $C(X) = C(Z, \tilde{W})$, there is no need to force $C(X_0) \subset C(\tilde{W})$. I have chosen this assumption because it is similar *in spirit* to the regularization method. There is no problem testing $\pi_1 = 0$ in this model. The $|t|$ statistic is the square root of the F statistic which is

$$F = \frac{(Y - Zg_0)'(I - M_0)Z[Z'(I - M_0)Z]^{-1}Z'(I - M_0)(Y - Zg_0)}{(Y - Zg_0)'\{I - M_0 - (I - M_0)Z[Z'(I - M_0)Z]^{-1}Z'(I - M_0)\}(Y - Zg_0)/[n - r(Z, X_0)]}.$$

where $M_0 \equiv X_0(X_0'X_0)^- X_0$ is the perpendicular projection operator onto $C(X_0)$. The test presupposes that you have some degrees of freedom for error in the reduced model. While $C[Z, X_0] \subset C(X)$, to get enough degrees of freedom to conduct a test you need $rank[Z, X_0] < rank(X)$.

Note that $(I - M_0)(Y - Zg_0) = Y - Zg_0 - X_0\hat{\pi}_3$ where $\hat{\pi}_3$ is the least squares estimate from the null model $Y - Zg_0 = X_0\pi_3 + e$.

An alternative idea for deriving a test of $\pi_1 = 0$, more similar to using the transformations based on known parameters discussed earlier, is to treat π_3 as known and test for $\pi_1 = 0$ in the linear model with one predictor variable,

$$\mathrm{E}(Y - Zg_0 - X_0\pi_3) = Z\pi_1.$$

With π_3 known, $Y - Zg_0 - X_0\pi_3$ is an observable dependent variable vector and the test is immediate. The $|t|$ statistic is the square root of the F statistic which is

$$F = \frac{(Y - Zg_0 - X_0\pi_3)'Z(Z'Z)^{-1}Z'(Y - Zg_0 - X_0\pi_3)}{(Y - Zg_0 - X_0\pi_3)'[I - Z(Z'Z)^{-1}Z'](Y - Zg_0 - X_0\pi_3)/(n - 1)}.$$

Of course π_3 is not known, so the obvious thing is to estimate it. The standard least squares estimate comes from fitting (4.3) using analysis of covariance ideas. ZB's approach of estimating

things under the null model is more similar to an older *ad hoc* test that comes from fitting a "stagewise regression," e.g. Alley (1987), Casella (1988), and Christensen (1988) and, for testing lack of fit, Neill and Johnson (1985). The stagewise approach is to test $\pi_1 = 0$ in

$$(Y - Zg_0 - X_0\hat{\pi}_3) = Z\pi_1 + e \tag{4.4}$$

where $\hat{\pi}_3$ is estimated from fitting

$$\mathrm{E}(Y - Zg_0) = X_0\pi_3 \tag{4.5}$$

with least squares. One nice thing about the stagewise approach is that, under model (4.3) and the hypothesis $\pi_1 = 0$, the estimate of π_3 is being obtained from a correct model (4.5) so it should have good asymptotic properties under the null hypothesis.

The stagewise estimate of π_1 is

$$\hat{\pi}_{1S} = [Z'Z]^{-1}Z'(Y - Zg_0 - X_0\hat{\pi}_3) = [Z'Z]^{-1}Z'(I - M_0)(Y - Zg_0)$$

whereas the least squares estimate of π_1 from model (4.3) is

$$\hat{\pi}_{1L} = [Z'(I - M_0)Z]^{-1}Z'(I - M_0)(Y - Zg_0) = [Z'(I - M_0)Z]^{-1}Z'(Y - Zg_0 - X_0\hat{\pi}_3)$$

The difference is that stagewise does not adjust the predictor variable Z for fitting X_0.

4.2 Regularized Linear Model Methods

Penalized least squares methods are valuable precisely because they give fitted values and predictions that are similar to fitting a reduced model. Regularization is used to obtain the advantages of fitting reduced models without actually specifying reduced models. Unfortunately, regularization does not provide as clear an approach to testing as fitting reduced models. Rather than explicitly choosing an X_0 we accomplish a similar result by using penalized estimates.

A stagewise approach to testing $\pi_1 = 0$ is to perform the test associated with fitting

$$(Y - Zg_0 - \tilde{W}\hat{\pi}_2) = Z\pi_1 + e \tag{4.6}$$

where $\hat{\pi}_2$ is estimated from fitting

$$\mathrm{E}(Y - Zg_0) = \tilde{W}\pi_2 \tag{4.7}$$

using penalized least squares. The assumptions necessary are that $\pi_1 = (a'\beta_* - g_0)$, the relatively weak assumptions necessary for getting a valid asymptotic test from (4.6) when $\hat{\pi}_2 = \pi_2$, and whatever assumptions are necessary for $\hat{\pi}_2$ to be consistent under penalized least squares and the null model.

An alternative to the stagewise approach would be to fit model (4.2) directly using penalized least squares where the penalty applies only to π_2 and perform an asymptotic test of $\pi_1 = 0$. For this, the *additional strong assumptions* are that $\pi_1 = a'\beta_* - g_0$, that $\pi_2 = U_a'\beta_*$, and that the true π_2 has a small value of the penalty function.

In estimation problems one can think of penalization as just a device to improve estimation. In this testing problem, you really need to believe that the true π_2 displays characteristics that are consistent with a small penalty function. It is not a great leap to assume that π_2 has a prior distribution determined by the penalty function and pursue a Bayesian test.

5 Conclusions

Regardless of the sizes of n and p, in a linear model

$$Y = X\beta_* + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma_\varepsilon^2 I$$

it is only possible to test a nonestimable function $a'\beta_*$ by making additional strong assumptions about the observable quantities and parameters.

References

Alley, W. M. (1987), A note on stagewise regression, *The American Statistician*, 41, 132-134.

Casella, G. (1988), Comparing regression coefficients across models, *The American Statistician*, 42, 91.

Christensen, Ronald (1988). Estimating the variance in stagewise regression. *The American Statistician*, 42, 288.

Christensen, Ronald (2001). *Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization*, Second Edition. Springer, New York.

Christensen, Ronald (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer, New York.

Cook, R. Dennis, Forzani, Liliana, and Rothman, Adam J. (2013). Prediction in abundant high-dimensional linear regression. *Electronic Journal of Statistics*, 7, 30593088.

Cook, R. Dennis, Forzani, Liliana, and Rothman, Adam J. (2015). Comment on Tarpey et al. (2014). *The American Statistician*, 69, 253-254.

Ding, P. (2014), Letter to the Editor on Tarpey et al. (2014), *The American Statistician*, 67, 316.

Ding, P. (2015). Reply. *The American Statistician*, 69, 255-256,

Fugate, Michael L., Christensen, Ronald, Hush, Don, and Scovel, Clint (2002). An equivalence relation between parallel calibration, Q matrix, and principal component regression. *Journal of Chemometrics*, 16(#1), 68-70.

Neill, J. W. and Johnson, D. E. (1985). Testing linear regression function adequacy without replication. *Annals of Statistics*, **13**, 1482-1489.

Tarpey, Thaddeus, Ogden, R., Petkova, Eva, and Christensen, Ronald (2015) Reply. *The American Statistician*, 69, 254-255.

Tarpey, Thaddeus, Ogden, R. Todd, Petkova, Eva, and Christensen, Ronald (2014). A Paradoxical Result in Estimating Regression Coefficients. *The American Statistician*, 68, 271-276.

Zhu, Yinchu and Bradic, Jelena (2017). Linear Hypothesis Testing in Dense High-Dimensional Linear Models. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2017.1356319