特集　経済計測の諸問題

Optimal Pre-Testing Procedure in Regression*

—— A minimum Average Risk Approach ——

Toshihisa Toyoda・Kazuhiro Ohtani

1

Although the use of pre-testing procedures in regression is so popular in applied econometric research, applied researchers have paid very little attention to the optimality or nonoptimality of their preliminary tests. Recently, some theoretical contributions have been made in setting some optimal critical values for pre-tests. Judge and his associates (e. g., [1] [2]) called our attention to optimal pre-testing problems on varied occasions. Sawa and Hiromatsu [4], based on their own specific risk function, proposed minimax regret critical values for pre-test estimators for the case of a single hypothesis. Brook [3], based on a more general quadratic risk function than the Sawa and Hiromatsu case, derived minimax regret critical values for multiple hypotheses. Toyoda and Wallace [5], based on the Brook type quadratic risk function, gave minimum average relative risk critical values. The minimax regret critical values for pre-tests which have been derived by Sawa and Hiromatsu [4] and Brook [3] are very stable and concentrated around $1.7\sim2.1$[1], while the alternative optimal critical values which have been derived by Toyoda and Wallace [5] are not necessarily close to the former values; they are very close to the minimax regret values when the number of restrictions is large but approach to the minimax (rather than the minimax regret) values when it is small.

In this paper we examine the Toyoda and Wallace approach when a reasonable skewed prior (rather than a diffuse one) of an unknown parameter is used and show how the optimal critical values are sensible to our prior knowledge of null hypothesis used in pre-tests.

The organization of the paper is as follows. In section 2 the basic pre-testing procedure is reviewed. In section 3 we show analytical solutions of optimal critical values when a gamma prior is used. In section 4 we state the numerical results of the optimal critical values for some specified values of the gamma parameters. Some concluding remarks are observed in the last section.

2

Let us consider a linear regression model

(1) $y=X\beta+\varepsilon\,;\ \varepsilon\sim N(0,\sigma^2I_T)$,

where y is, say, $T\times1$, X is $T\times k$, β is $k\times1$ and ε is $T\times1$. We assume that X is a known matrix with rank k.

We assume that (1) is an incompletely specified model and that we conduct some kind of preliminary tests prior to estimating β or $E(y|X)$. Various kinds of hypothesis for pre-tests can be expressed conveniently by

(2) $H'\beta=h$,

where H' is $m\times k$ with rank m and h is $m\times1$. Both H' and h are assumed to be known. Null hypotheses of insignificance of any subset of explanatory variables and of linear dependence among them are of course expressed by (2). In order to show the wide applicability of the general linear hypothesis (2) to incompletely specified models, let us consider the following two examples which are used frequently in applied research.

First, consider pooling T time series of N_t cross-sections in regression analysis ($t=1, 2, \cdots, T$). The model is specified as

(3) $y_t=X_t\beta_t+\varepsilon_t\,;\ \varepsilon_t\sim N(0,\sigma^2I_{N_t})$,

where N_t observations of y_t and X_t are available in each period ($t=1, 2, \cdots, T$). We decide whether we pool the whole time series of cross-sections or not, based on the outcome of a test of the hypothesis

* Many helpful discussions with Professor T. Dudley Wallace are greatly acknowledged. We also thank Professor Keiichiro Nakayama and the referee of this journal for helpful comments on an earlier version of this article.

1) This does not mean that the levels of significance corresponding to the optimal critical values are stable.

(4) $H_0:\beta_1=\beta_2=\cdots=\beta_T$.

In this case, define

$$H'=\begin{pmatrix} 1 & -1 & 0\cdots\cdots 0 & 0 \\ 0 & 1 & -1\cdots\cdots 0 & 0 \\ \vdots & \vdots & \vdots\ \ddots\quad \vdots & \vdots \\ 0 & 0 & 0\cdots\cdots 1 & -1 \end{pmatrix}\otimes I_k,$$
$$(T-1)\times T$$

$$\beta'=(\beta_{11},\beta_{12},\cdots,\beta_{1k},\cdots,\beta_{T1},\beta_{T2},\cdots,\beta_{Tk}),$$

and

$$h=0,$$

and the hypothesis (4) can be expressed by (2).

The second example is concerned with the Chow test, a test of equality of regression coefficients across two data regimes. Wallace [7] points out that this problem is equivalent to choosing H' to be $(I_k, -I_k)$.

Let us consider a pre-test estimator for $E(y|X)$[2]:

(5) $X\beta^*=\begin{cases} Xb & \text{if } u\geqq\lambda \\ X\hat{\beta} & \text{if } u<\lambda \end{cases}$

where b is the unrestricted least squares estimator for β, $\hat{\beta}$ is the restricted least squares estimator, λ is a critical value for pre-testing of the hypothesis (2) and u is the calculated value of a test statistic. That is,

(6) $u=(H'b-h)'(H'S^{-1}H)^{-1}(H'b-h)/m\hat{\sigma}^2$

where $S=X'X$ and $\hat{\sigma}^2$ is an unbiased estimator of σ^2, and u is distributed as the noncentral F with m and T–k degrees of freedom and noncentrality θ, where

(7) $\theta=(H'\beta-h)'(H'S^{-1}H)^{-1}(H'\beta-h)/2\sigma^2$.

A quadratic risk function for the pre-test estimator is defined as[3]

(8) $R(X\beta^*)=E(X\beta^*-X\beta)'(X\beta^*-X\beta)$.

After some calculations it is found that[4]

(9) $R(X\beta^*)=\sigma^2\{k-m+mr(\theta,\lambda)$
$\qquad+2\theta[1-2r(\theta,\lambda)+s(\theta,\lambda)]\}$,

where

(10) $r(\theta,\lambda)=Pr\{F'(m+2, T-k;\theta)\geqq m\lambda/(m+2)\}$

and

(11) $s(\theta,\lambda)=Pr(F'(m+4, T-k;\theta)\geqq m\lambda/(m+4)\}$.

Here, $F'(a,b;\theta)$ stands for a statistic which has the noncentral F density with a and b degrees of freedom and noncentrality θ. From (9) it is seen that the risk function divided by σ^2 depends on m, k, T, λ and θ.

2) Here, we do not consider a pre-test estimator for β *per se*, whose general quadratic risk function is not scale invariant to measuring variables in X. Investigating any optimal pre-testing procedure for this case is an open question, although Brook [3] has given some limited examples of minimax regret critical values.

3) This is the same type risk function as has been used by Brook [3]. If we deal with orthogonal data, this risk function is equivalent to the one of the pre-test estimator for β *per se* because it holds that

$$E(X\beta^*-X\beta)'(X\beta^*-X\beta)$$
$$=E(\beta^*-\beta)'S(\beta^*-\beta)=E(\beta^*-\beta)'(\beta^*-\beta).$$

4) See Brook [3] or Toyoda and Wallace [5].

3

Toyoda and Wallace [5] seeked the values of λ which minimized

(12) $G(\lambda)=\int_0^\infty\{R(X\beta^*)-\min[R(Xb),R(X\hat{\beta})]\}d\theta/\sigma^2$.

If a diffuse prior on θ is assumed, minimizing $G(\lambda)$ in (12) is equivalent to minimizing expected relative risk. The approach adopted by Toyoda and Wallace is very natural and general under the ignorance of prior distribution of θ. Logically, θ can take any value in the interval $[0,\infty)$, depending on how we set the null hypothesis for the pre-tests. However, on many occasions some skewed distributions of θ will be more proper than the diffuse one[5].

In the following we use a gamma prior on θ, which is a generally accepted representative of skewed distributions on the domain $[0,\infty)$, in order to check the sensitivity of the optimal critical values obtained by Toyoda and Wallace.

Let the gamma distribution be

(13) $g(\theta;\mu,\alpha)\propto\alpha\theta^{\mu-1}e^{-\alpha\theta}$; $\quad\mu>0, \alpha>0$.

We first note that in the present case the Toyoda and Wallace criterion can be simplified to a minimum average risk criterion from the former minimum average *relative* risk. Toyoda and Wallace used the *relative* risk so that $G(\lambda)$ could be integrable. However, our new average risk incorporating the gamma prior is integrable by itself. Its explicit from is:

5) Consider, for instance, estimating an aggregate consumption function: $C_t=\beta_0+\beta_1Y_t^*+\beta_2r_t+u_t$, where C=consumption expenditures, Y^*=permanent income, and r=the rate of interest. Our a priori knowledge suggests that the prior distribution of θ for the case of testing $H:\beta_1=0$ would be skewed for some positive value of θ while the one for the case of testing $H:\beta_2=0$ would be skewed for $\theta=0$.

$$(14)\quad E_\theta[R(X\beta^*)] \propto \int_0^\infty \{k-m+mr(\theta,\lambda) +2\theta[1-2r(\theta,\lambda)+s(\theta,\lambda)]\}\theta^{\mu-1}e^{-\alpha\theta}d\theta$$
$$= -m\int_0^\infty \theta^{\mu-1}e^{-\alpha\theta}[1-r(\theta,\lambda)]d\theta +2\int_0^\infty \theta^{\mu}e^{-\alpha\theta}\{2[1-r(\theta,\lambda)] -[1-s(\theta,\lambda)]\}d\theta +k\int_0^\infty \theta^{\mu-1}e^{-\alpha\theta}d\theta$$

Let $f(F';m^*,T-k,\theta)$ be $\frac{m^*}{T-k}f'(F';m^*,T-k,\theta)$ where f' is the noncentral F density with m^*, $T-k$ degrees of freedom and noncentrality θ, and ${}_2F_1(.,.;.;.)$ represent the Gauss hypergeometric function. Then, the necessary condition for minimizing $E_\theta[R(X\beta^*)]$ is the following.

$$(15)\quad \frac{dE_\theta[R(X\beta^*)]}{d\lambda} \propto -m\int_0^\infty \theta^{\mu-1}e^{-\alpha\theta}f\left(\frac{m\lambda}{T-k};m+2,T-k,\theta\right)d\theta +4\int_0^\infty \theta^{\mu}e^{-\alpha\theta}f\left(\frac{m\lambda}{T-k};m+2,T-k,\theta\right)d\theta -2\int_0^\infty \theta^{\mu}e^{-\alpha\theta}f\left(\frac{m\lambda}{T-k};m+4,T-k,\theta\right)d\theta$$
$$\propto \frac{\lambda^{m/2}}{(T-k+m\lambda)^{(m+T-k)/2+1}} \times\Bigg[-m\,{}_2F_1\left(\mu,\frac{m+T-k}{2}+1,\frac{m}{2}+1;\frac{m\lambda}{(\alpha+1)(m\lambda+T-k)}\right) +\frac{4\mu}{\alpha+1}{}_2F_1\left(\mu+1,\frac{m+T-k}{2}+1;\frac{m}{2}+1;\frac{m\lambda}{(\alpha+1)(m\lambda+T-k)}\right) -\frac{2m(m+T-k+2)\lambda\mu}{(m+2)(m\lambda+T-k)(\alpha+1)}{}_2F_1\left(\mu+1,\frac{m+T-k}{2}+2;\frac{m}{2}+2;\frac{m\lambda}{(\alpha+1)(m\lambda+T-k)}\right)\Bigg]$$
$$=0.$$

That is, the minimum is attained when

$(16)\quad \lambda=0$

or

$(17)\quad [\ \ ]=0.$

It is shown that the inside of the bracket of (17) reduces to

$$(18)\quad y=(\alpha m+m-4\mu)\,{}_2F_1\left(\mu,\frac{m+T-k}{2}+1;\frac{m}{2}+1;y\right)\Big/\left\{2(1-\alpha)d\left[{}_2F_1\left(\mu,\frac{m+T-k}{2}+1;\frac{m}{2}+1;y\right]\Big/dy\right\}$$

where

$(19)\quad y=m\lambda/(\alpha+1)(m\lambda+T-k)$ [6]

provided that $\lambda\neq 0$ and $(1-\alpha)(\alpha m+m-4\mu)>0$. It is also seen that the optimal value of λ for the case $(1-\alpha)(\alpha m+m-4\mu)<0$ is

$(20)\quad \lambda=0\quad$ (i. e., $y=0$).

Proofs for these results are given in Appendix.

4

Comparing the fixed point solution (18) with the one obtained by Toyoda and Wallace [5] reveals that they become coincident when $\alpha=0$ and $\mu=1$; it is a natural result since the gamma prior becomes uniform in this case. Therefore, equation (18) generalizes the Toyoda and Wallace result in one sense, i. e., from the diffuse to the non-diffuse priors on θ.

However, it is not always possible to transform the right hand side of (18) into a function only of the incomplete beta function ratios as Toyoda and Wallace could manage. Then, we have developed a computer program of the Gauss hypergeometric series to get the fixed point solutions of equation (18) directly. We have used an iterative search procedure to compute them, differing from zero, to three decimal places for various selected values of the parameters (α and μ) and degrees of freedom (m and $T-k$)[7].

According to our numerical results, it is seen that $\partial\lambda^*/\partial\alpha>0$ and $\partial\lambda^*/\partial\mu<0$, i. e., the optimal critical values behave monotonically but to the opposite ways for changes in the parameters, α and μ, given any fixed numerator and denominator degrees of freedom. If α is comparatively small (say, .05), the potimal values take fairly stable points around 1.0~4.0 depending on the degrees of freedom. However, if α is comparatively large (say, .5), the optimal values become very sensitive to the value of μ, particularly for small μ. A typical example of the behavior of λ^*

---

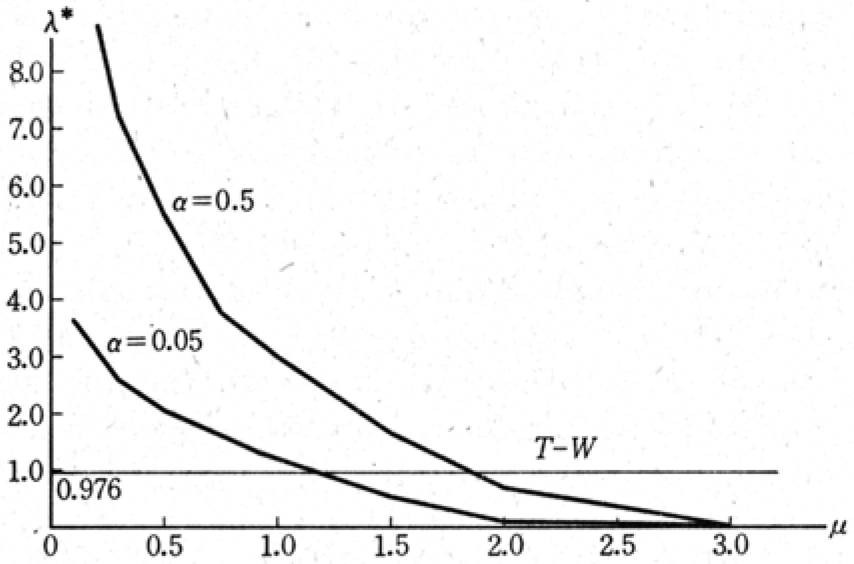
6) As $\lambda\geqq 0$ and $\lambda=\frac{(\alpha+1)(T-k)y}{m(1-(\alpha+1)y)}$, it should hold that $0\leqq y<\frac{1}{\alpha+1}$.

7) We have selected .05, .1, .3 and .5 for the values of α and .1, .3, .5, .8, .9, 1.0., 1.1, 1.2, 1.5, 2.0 and 3.0 for the ones of μ. From the condition $\lambda=0$, it must hold that $\alpha>1 \Rightarrow m<4\mu/(1+\alpha)$ or $\alpha<1 \Rightarrow m>4\mu/(1+\alpha)$. Considering these conditions, we have concentrated our numerical analysis on the cases $m=8$, 16 and 24.

for various values of α and μ is exhibited in Figure 1, where $T-W$ stands for the optimal critical value given in Toyoda and Wallace [5] which is .976.

Now, let us compare the present optimal critical values, λ^*, with λ' which have been found by Toyoda and Wallace [5]. When $\alpha=.05$ and μ is around 1.0, i. e., when the investigator has relatively weak prior knowledge about the hypothesis, the values of λ^* are nearly equal to the ones of λ'. That is to say, the minimum average relative risk critical values roughly correspond to the minimum average risk critical values when the gamma prior with $\alpha=.05$ and $\mu\simeq 1.0$ is used. This is a natural result and also confirms the Toyoda and Wallace result because the gamma distribution with such parameter values is very flat. When $\alpha=.5$ and particularly when $\mu<1.0$, i. e., when the investigator has relatively firm prior knowledge about the hypothesis, the values of λ^* are far greater than the ones of λ'.

Figure 1 Behavior of λ^* for various values of the parameters and for $m=8$, $T-k=16$



5

In performing preliminary tests in regression analysis, the investigator sometimes has prior knowledge of plausibility or implausibility of the hypothesis. The degree of prior confidence can be expressed by a parameter (or parameters) of a skewed prior distribution of the noncentrality parameter, θ. We have adopted the gamma prior which seems to be reasonable in this context. We have first shown that the minimum average risk criterion (rather than the minimum average relative risk one) can be used in order to find optimal critical values for pre-tests.

To obtain some specific optimal critical values, we have selected some values for the gamma parameters and the ones for the numerator and denominator degrees of freedom. We have found that when the distributional parameters take such values as the gamma prior distribution becomes very flat, the present optimal critical values are almost equal to the ones found by Toyoda and Wallace [5].

However, when α and μ are assigned such values as the gamma prior skews for relatively small (large) values of θ, the optimal critical values increase (decrease) considerably. It should be noted that the optimal critical values are not robust with respect to the both numerator and denominator degrees of freedom just as Toyoda and Wallace [5] showed.

Given no prior knowledge of the plausibility of the hypothesis, the Toyoda and Wallace approach [5] which utilizes a diffuse prior would be a reasonable strategy. However, when one has some prior confidence of the plausibility (or implausibility) of the hypothesis, it has become evident that he should use some larger (smaller) critical values and therefore the corresponding smaller (larger) significance levels for pre-tests than the Toyoda and Wallace values. In any case, the conventional pre-testing procedure using 1 or 5 per cent significance level would be misleading.

(Toshihisa Toyoda : Kobe University)
(Kazuhiro Ohtani : Kobe University of Commerce)

Appendix

First, we note the following relation:

$$ {}_2F_1(a+1, b; c; y) - {}_2F_1(a, b; c; y) = \frac{by}{c}\, {}_2F_1(a+1, b+1; c; y), $$

where

$$ {}_2F_1(a, b; c; y) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{i=0}^{\infty} \frac{\Gamma(a+i)\Gamma(b+i)}{\Gamma(c+i)} \frac{y^i}{i!}. $$

Combining the above relation with the differentiation formula (e. g., [6], p. 557)

$$ d[{}_2F_1(a, b; c; y)]/dy = \frac{ab}{c}\, {}_2F_1(a+1, b+1; c+1; y), $$

the inside of the bracket of (17) reduces to

$$ m\, {}_2F_1\left(\mu, \frac{m+T-k}{2}+1; \frac{m}{2}+1; y\right) - \left(\frac{4\mu}{\alpha+1}\right)\Bigg\{ {}_2F_1\left(\mu, \frac{m+T-k}{2}+1; \frac{m}{2}+1; y\right) $$

$$+\frac{y}{\mu}d\left[{}_2F_1\left(\mu,\frac{m+T-k}{2}+1;\frac{m}{2}+1;y\right)\right]$$

$$\Big/dy\Big\}+\frac{2(m+T-k+2)\left(\frac{m}{2}+1\right)y}{(m+2)\left(\frac{m+T-k}{2}+1\right)}$$

$$d\left[{}_2F_1\left(\mu,\frac{m+T-k}{2}+1;\frac{m}{2}+1;y\right)\right]\Big/dy$$

$$=0.$$

Rearranging the result, we obtain the equation (18).

If $(1-\alpha)(\alpha m+m-4\mu)<0$, the right hand side of (18) is negative. However, y is always non-negative from (19), which is a contradiction. Therefore, [ ] is not zero. Hence, (20) holds because of (15).

References

[1] Bock, M. E., T. A. Yancey and G. G. Judge, "The Statistical Consequences of Preliminary Test Estimators in Regression," *Journal of the American Statistical Association*, Vol. 68, No. 341 (March 1973), pp. 109–16.

[2] Bock ,M. E., G. G. Judge and T. A. Yancey, "Some Comments on Estimation in Regression after Preliminary Tests of Significance," *Journal of Econometrics*, Vol. 1, No. 2 (May 1973), pp. 191–200.

[3] Brook, R. J., "On the Use of a Regret Function to Set Significance Points in Prior Tests of Estimation," *Journal of the American Statistical Association*, Vol. 71, No. 353 (March 1976), pp. 126–31.

[4] Sawa, T. and T. Hiromatsu, "Minimax Regret Significance Points for a Preliminary Test in Regression Analysis," *Econometrica*, Vol. 41, No. 6 (November 1973), pp. 1093–1102.

[5] Toyoda, T. and T. D. Wallace, "Optimal Critical Values for Pre-Testing in Regression," *Econometrica*, Vol. 44, No. 2 (March 1976), pp. 365–75.

[6] U. S. Department of Commerce, National Bureau of Standards, *Handbook of Mathematical Functions*. Applied Mathematics Series, Vol. 55, Washington, D. C. (1964).

[7] Wallace, T. D., "Weaker Criteria and Tests for Linear Restrictions in Regression," *Econometrica*, Vol. 40, No. 4 (July 1972), pp. 689–98.