

## THE TRADITIONAL PRETEST ESTIMATOR\*

J. R. MAGNUS†

**Abstract.** We consider the problem of estimating  $k$  coefficients of interest in a linear regression model when the  $(k + 1)$ st coefficient is of no interest. The traditional pretest estimator is a two-step estimator of the coefficients of interest based on a  $t$ -test for the  $(k + 1)$ st coefficient. We study the behavior of this estimator. Questions of admissibility, risk, and regret are addressed.

**Key words.** regression analysis, model selection, biased estimation, univariate normal mean, mean squared error criterion, minimax regret

PII. S0040585X97977604

**1. Introduction.** The purpose of this paper is to study the traditional pretest estimator in the context of developing an optimal method of estimation under model uncertainty. Our starting point is the linear regression model

$$y = X\beta + Z\gamma + u, \quad u \sim N(0, \sigma^2 I_n),$$

where  $X$  is an  $n \times k$  matrix of explanatory variables that are required to be in the model on theoretical or other grounds, while  $Z$  is an  $n \times m$  matrix of additional explanatory variables about which there is doubt as to whether they should be in the model. (In Leamer's [29], [30] terminology,  $X$  contains the "free" variables and  $Z$  the "doubtful" variables.) We are interested in estimating  $\beta$  (or specified linear combinations of its elements), while the value of  $\gamma$  is of no interest to us. Thus,  $\beta$  is a vector of parameters of interest and  $\gamma$  is a vector of nuisance parameters. The only reason for including  $Z$  in the model is that by doing so we expect to obtain a "better" estimator of  $\beta$ . In this context we assess the relative performance of estimators by the mean squared error (MSE) criterion. Many of the results in this paper do not require the normality assumption. We shall point these out as we go along.

The most common approach is to test the hypothesis that  $\gamma = 0$  and to include  $Z$  if the hypothesis is rejected and exclude it otherwise. Inference on parameter estimators is then carried out as if the resulting model were correct. Inferences made in this way will, however, be invalid since the model selection procedure influences the properties of the estimator. This paper is concerned with determining the implications of model selection on the estimators of the parameters of interest. This problem was already discussed following Tinbergen's [37] study for the League of Nations. Both Keynes [26] and Friedman [14], in their respective critiques of Tinbergen's work, focused on the method of model selection when the estimation procedure repeatedly uses the same data to discriminate among plausible competing theories. The same point was made in Haavelmo's seminal paper [18, section 17]. Koopmans [27] suggested that a completely new theory of inference was required to solve the dilemmas implied by the model selection problem.

---

\*Received by the editors September 7, 1998.

<http://www.siam.org/journals/tvp/44-2/97760.html>

†CentER, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, The Netherlands (Magnus@kub.nl).

The theory of pretest estimation is an attempt to address this problem, but, as we shall see, is not a successful one. Early work on the traditional pretest estimator includes [1], [2], [20], [28], [10], [42], [36], [7], and [6]. See the surveys [21], [22], [24], and [15]. For a Bayesian perspective see [44], [31], and [16].

For simplicity we assume that  $Z$  contains a single explanatory variable ( $m = 1$ ) and that the disturbance variance  $\sigma^2$  is known. The problem under consideration is, therefore, to estimate  $k$  coefficients of interest (or linear combinations of these) in a regression model containing  $k + 1$  regressors using an MSE criterion in a situation where the value of the  $(k + 1)$ st coefficient is of no interest. We call this the *regression problem*. Remarkably, it turns out that this problem is equivalent to another problem which appears to be quite different, namely, the problem of finding the “best” estimator of  $\theta$  when we have one observation  $x$  from an  $N(\theta, 1)$  distribution. We call this the  $N(\theta, 1)$  *problem*. The equivalence between the regression problem and the  $N(\theta, 1)$  problem (the equivalence theorem) was proved, in a more general context, by Magnus and Durbin [32].

Let  $b_r$  be the least squares (LS) estimator of  $\beta$  assuming that  $\gamma = 0$ , let  $b_u$  be the LS estimator of  $\beta$  assuming that  $\gamma$  is also estimated, and let  $\theta$  be the value of  $\gamma$  divided by the standard error of its LS estimator. In the traditional approach to the regression problem a choice is made between  $b_r$  and  $b_u$  depending on the outcome of a test of the hypothesis  $\theta = 0$ . The traditional pretest estimator thus has the form  $b = b_r$  if  $|\hat{\gamma}| \leq c(\text{SE}(\hat{\gamma}))$  and  $b = b_u$  if  $|\hat{\gamma}| > c(\text{SE}(\hat{\gamma}))$ , where  $\hat{\gamma}$  is the LS estimator of  $\gamma$ ,  $\text{SE}(\hat{\gamma})$  is its estimated standard error, and  $c > 0$ . The equivalent pretest estimator for the  $N(\theta, 1)$  problem is  $t = 0$  if  $|x| \leq c$  and  $t = x$  if  $|x| > c$ . We demonstrate a number of undesirable properties of this estimator. First, it is inadmissible. Second, there is a range of values of  $\theta$  for which the MSE of  $t$  is greater than the MSE of both the “usual” estimator  $t = x$  and the “silly” estimator  $t = 0$ . Third, in the neighborhood of the value  $|\theta| = 1$ , which we show to be of crucial significance at various points of the paper, the MSE is maximized when  $c = 1.91$ . This means that when the traditional pretest is carried out at the usual 5% level, the resulting estimator is close to having worst possible performance.

In Figure 1 we have graphed four estimators  $t(x)$  of  $\theta$  when  $x \sim N(\theta, 1)$ : the “silly” estimator  $t(x) = 0$  (which corresponds to the restricted estimator  $b_r$ ), the “usual” estimator  $t(x) = x$  (which corresponds to  $b_u$ ), the “pretest” estimator  $t(x) = 0$  if  $|x| \leq c$ ,  $t(x) = x$  if  $|x| > c$  (which corresponds to the traditional pretest estimator), and the “Laplace” estimator, our own favorite. It is clear from Figure 1 that the pretest estimator cannot have good properties, due to its discontinuity. Because of the equivalence theorem, the regression problem reduces to the following: Given a single observation from an  $N(\theta, 1)$  distribution, find the best estimator of  $\theta$ . This seemingly trivial problem is far from trivial. It was considered in detail by Magnus [33]. A wide range of possible estimators was considered, the one finally preferred having an interpretation as a Bayes estimator with prior density  $\frac{1}{2} c \exp(-c|\theta|)$ ,  $-\infty < \theta < \infty$ ,  $c = \log 2$ . This is our “Laplace” estimator.

The plan of this paper is as follows. In section 2 we define the problem, introduce the relevant notation, and present the equivalence theorem. In section 3 the inappropriateness of the pretest estimator is presented from a different perspective. Sections 4 and 5 contain the principal results in terms of risk (section 4) and regret (section 5). Section 6 offers some conclusions, and section 7 contains the proofs.

**2. Equivalence of the regression problem and the  $N(\theta, 1)$  problem.** In the regression problem we are concerned with the estimation of (linear combinations

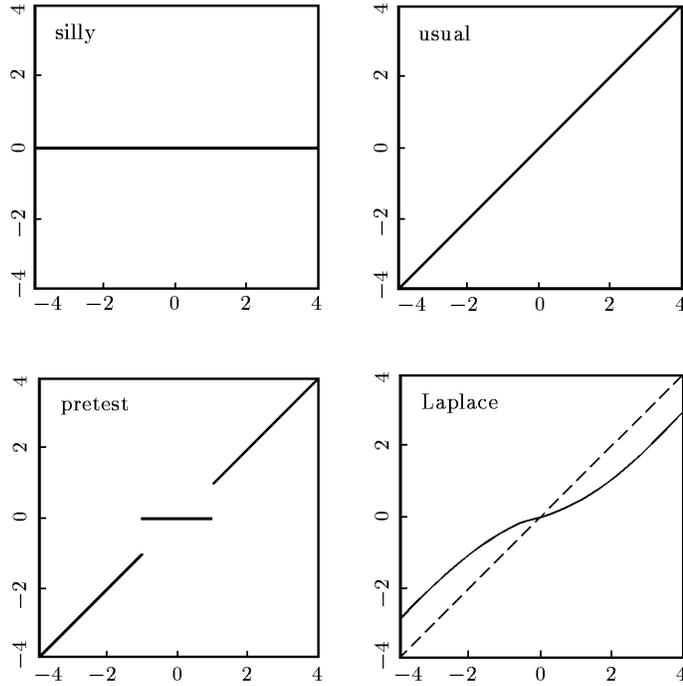


FIG. 1. Four estimators  $t(x)$  of  $\theta$  when  $x \sim N(\theta, 1)$ .

of the elements of  $\beta$  in the linear regression model  $y = X\beta + \gamma z + u$ , where  $y(n \times 1)$  is the vector of observations;  $X(n \times k)$  and  $z(n \times 1)$ , both nonrandom, represent the regressors;  $u(n \times 1)$  is a random vector of unobservable disturbances; and  $\beta(k \times 1)$  and the scalar  $\gamma$  are unknown nonrandom parameters. We assume that the design matrix  $(X: z)$  has full column-rank and that the disturbances  $u_1, \dots, u_n$  are i.i.d.  $N(0, \sigma^2)$ , that is,  $u \sim N(0, \sigma^2 I_n)$ ,  $\sigma^2 > 0$ .

We assume also that  $\sigma^2$  is known. This is, of course, unrealistic but it simplifies the analysis without affecting the main results.

We introduce the following notation:

$$M = I_n - X(X'X)^{-1}X', \quad q = \frac{\sigma}{\sqrt{z'Mz}} (X'X)^{-1}X'z, \quad \theta = \frac{\gamma}{\sigma/\sqrt{z'Mz}}.$$

(Notice that the rank condition implies that  $z'Mz > 0$ .) The idempotent matrix  $M$  and the vector  $q$  are known, while  $\theta$  is unknown (since  $\gamma$  is unknown). The parameter  $\theta$  plays an important role in our analysis and we shall call it the *theoretical t-ratio*.

The LS (maximum likelihood) estimators of  $\beta$  and  $\gamma$  are

$$(2.1) \quad b_u = b_r - \hat{\theta}q, \quad \hat{\gamma} = \frac{z'My}{z'Mz},$$

where

$$(2.2) \quad b_r = (X'X)^{-1}X'y,$$

$$\hat{\theta} = \frac{\hat{\gamma}}{\sigma/\sqrt{z'Mz}}.$$

The subscripts “ $u$ ” and “ $r$ ” stand for “unrestricted” and “restricted” (with  $\gamma = 0$ ), respectively. It is clear that  $\hat{\theta} \sim N(\theta, 1)$ . We emphasize again that our interest lies in the estimation of (linear combinations of the elements of)  $\beta$ , the parameters of interest, while  $\gamma$  is essentially a nuisance parameter.

The traditional approach, used by the large majority of applied econometricians and statisticians, is to consider only two estimators of  $\beta$ : the restricted estimator  $b_r$  defined in (2.2) (where  $\gamma = 0$ ) and the unrestricted estimator  $b_u$  defined in (2.1). The choice between them is then based on the  $t$ -ratio  $\hat{\theta}$  and the estimator  $b$  for  $\beta$  can thus be written as

$$(2.3) \quad b = \begin{cases} b_r & \text{if } |\hat{\theta}| \leq c, \\ b_u & \text{if } |\hat{\theta}| > c, \end{cases}$$

for some  $c \geq 0$ . For example,  $c = 1.96$  and  $c = 2.58$  correspond to the 5% and the 1% significance levels, respectively. Hence, the estimator employed in the traditional approach is (2.3), even though the large majority of investigators acts as if the estimator is either  $b_r$  or  $b_u$ .

Given that we are not interested in  $\gamma$  but only in the best possible estimation of  $\beta$ , this procedure makes very little sense. We are testing  $H_0: \theta = 0$  against  $H_1: \theta \neq 0$  (or equivalently,  $\gamma = 0$  against  $\gamma \neq 0$ ) and this gives us an answer to the question, Is it true that  $\theta = 0$ ? But this is the wrong question in this context. The right question is, Is  $b_r$  a better estimator for  $\beta$  than  $b_u$ ? If we agree to judge an estimator by its MSE,

$$\text{MSE}(b) \equiv \mathbf{E}(b - \beta)(b - \beta)' = \mathbf{D}(b) + (\mathbf{E}b - \beta)(\mathbf{E}b - \beta)',$$

then the following theorem, which is a special case of Theorem 1 in [32], will help us to answer this question.

**THEOREM 2.1.** *We have  $\text{MSE}(b_r) - \text{MSE}(b_u) = (\theta^2 - 1)qq'$  and hence*

$$\begin{aligned} \text{MSE}(b_r) &\leq \text{MSE}(b_u) && \text{if } \theta^2 < 1, \\ \text{MSE}(b_r) &= \text{MSE}(b_u) && \text{if } \theta^2 = 1, \\ \text{MSE}(b_r) &\geq \text{MSE}(b_u) && \text{if } \theta^2 > 1. \end{aligned}$$

Theorem 2.1 shows, still assuming that  $b_r$  and  $b_u$  are the only two estimators to choose between, that the choice should be based on the null hypothesis  $H_0: \theta^2 \leq 1$  against the alternative  $H_1: \theta^2 > 1$ . Toro-Vizcarrondo and Wallace [38] made this point and they showed that a uniformly most powerful test is obtained from the probability

$$\mathbf{P} \left\{ |\hat{\theta}| \leq c \mid \theta^2 = 1 \right\} = 1 - \alpha,$$

where  $\alpha$  denotes the level of significance. (See also [1], [2], [40], [43], [41], [17], [42], and [13].) For example, a 5% test and a 1% test now correspond to  $c = 2.65$  and  $c = 3.33$ , respectively. Application of the second test obviously leads to more frequent use of the restricted estimator.

We notice that both the first and the second test lead to an estimator  $b$  of  $\beta$  of the form (2.3). There are two problems in applying either of these two tests. The first is that the choice of significance level is largely arbitrary. The second problem is that, after the preliminary test in which we decide whether to use  $b_r$  or  $b_u$ , neither

of these two is the estimator of  $\beta$  actually used. The estimator actually used is given in (2.3) and is known as the *traditional pretest estimator*. Let us rewrite (2.3) as

$$(2.4) \quad b = \lambda(\hat{\theta}) b_u + (1 - \lambda(\hat{\theta})) b_r,$$

where  $\lambda(\hat{\theta}) = 0$  if  $|\hat{\theta}| \leq c$ ,  $\lambda(\hat{\theta}) = 1$  if  $|\hat{\theta}| > c$ . This formulation shows  $b$  as a weighted average of  $b_u$  and  $b_r$ . Any estimator of  $\beta$  of the form (2.4), where  $\lambda$  is a real-valued function of  $\hat{\theta}$  satisfying certain regularity conditions, will be called a *weighted-average least squares* (WALS) estimator. The pretest estimator is a very simple (and not a very good) example of such an estimator. It is intuitively appealing to think of a WALS estimator as a continuous function of  $\hat{\theta}$  such that the larger  $|\hat{\theta}|$  is, the more weight is given to  $b_u$  relative to  $b_r$ . WALS estimation can thus be viewed as a two-step regression procedure, where we first obtain  $\hat{\theta}$  and then  $b$  as a function of  $\hat{\theta}$ . The MSE of the WALS estimator  $b$ , defined in (2.4), is given in Theorem 2.2, which is a special case of Theorem 2 in [32].

**THEOREM 2.2** (the equivalence theorem). *The MSE of the WALS estimator  $b = \lambda b_u + (1 - \lambda) b_r$ , where  $\lambda = \lambda(\hat{\theta})$ , is given by*

$$\text{MSE}(b) = \sigma^2 (X'X)^{-1} + \text{MSE}(\tilde{\theta}) qq',$$

where  $\tilde{\theta} = \lambda \hat{\theta}$ .

We notice that neither Theorem 2.1 nor Theorem 2.2 depends on the normality assumption. We need only assume that the elements of  $u$  are uncorrelated  $(0, \sigma^2)$  variables.

In [33] we tried to find an optimal weighting function  $\lambda$  such that the WALS estimator  $b$  is best in the sense of having lowest MSE. One glance at Theorem 2.2 reveals that  $\text{MSE}(b)$  is minimized if and only if  $\text{MSE}(\tilde{\theta})$  is minimized. The striking consequence of Theorem 2.2 is therefore that finding the best WALS estimator of  $\beta$  is equivalent to finding the best estimator  $\tilde{\theta}$  of  $\theta$ .

In this paper we study only the traditional pretest estimator. Theorem 2.2 shows that studying the pretest estimator  $b$ , given in (2.3), for the regression problem is equivalent to studying the estimator  $\tilde{\theta}$ , given by

$$(2.5) \quad \tilde{\theta} = \begin{cases} 0 & \text{if } |\hat{\theta}| \leq c, \\ \hat{\theta} & \text{if } |\hat{\theta}| > c, \end{cases}$$

for the  $N(\theta, 1)$  problem. Thus, in sections 4 and 5 we consider a single observation  $\hat{\theta}$  from an  $N(\theta, 1)$  distribution and study the properties of  $\tilde{\theta}$  given in (2.5).

**3. A story.** The following story illustrates the inappropriateness of the pretest estimator.

*On the tiny remote island of I, the I-landers lived mainly on fish. Since the wind around the island varied from day to day, the I-landers had built two boats for their fishermen, the R-boat and the U-boat. The R-boat (R for "rest") was ideal when there was no wind, the U-boat (U for "unrest") in a heavy storm. Each evening after dinner the King announced his weather forecast for the next day upon which one of the two boats was prepared. An incorrect forecast of the weather and hence a wrong choice of boat could have serious consequences. All I-landers remembered the recent hurricane that the King had failed to forecast and in which the R-boat capsized, resulting in the death of all fishermen on board. One day a young adventurer A found himself*

stranded at I. A inspected the two fishing boats and found them well built for their purpose. He noticed, however, that the extreme weather conditions for which the boats were designed rarely occurred. Most days at the island saw a moderate breeze. A decided to build a boat himself. After several months, his work completed, he proudly presented his new boat to the assembled I-landers. "How does your boat perform when there is no wind?" asked one of the fishermen. "Well," said A, "you cannot expect my boat to do quite so well as your R-boat, which was built for that purpose, but it does definitely better than the U-boat when there is little or no wind." "And how does your boat perform in stormy weather?" asked a second fisherman. "Again," answered A, "your U-boat does better in a storm, but my boat performs better than the R-boat. In particular, my boat will not capsize in a storm." Then the King said: "If the weather is fair, with a gentle breeze, what performance has your boat?" Somewhat embarrassed A replied: "I must admit that under such conditions my boat performs worse than both the R-boat and the U-boat." "Throw this man into the ocean!" cried the King, and A was never heard of again.

**4. Risk.** The traditional pretest estimator for the regression problem is defined in (2.3):

$$(4.1) \quad b = \begin{cases} b_r & \text{if } |\hat{\theta}| \leq c, \\ b_u & \text{if } |\hat{\theta}| > c. \end{cases}$$

The equivalent estimator for the  $N(\theta, 1)$  problem is

$$(4.2) \quad t(x; c) = \lambda_c(x) x = \begin{cases} 0 & \text{if } |x| \leq c, \\ x & \text{if } |x| > c, \end{cases}$$

where

$$(4.3) \quad \lambda_c(x) = \begin{cases} 0 & \text{if } |x| \leq c, \\ 1 & \text{if } |x| > c. \end{cases}$$

We shall refer to the estimator (4.2) as the pretest estimator as well, if there is no possible ambiguity. We notice that  $\lambda_c$  is well defined at  $c = 0$  and at  $c = \infty$ . At  $c = 0$  we have  $\lambda_0 \equiv 1$  and  $t(x; 0) = x$ , the "usual" estimator. At  $c = \infty$  we have  $\lambda_\infty \equiv 0$  and  $t(x; \infty) = 0$ , the "silly" estimator.

The pretest estimator has a discontinuity at  $x = \pm c$  and we would therefore expect the estimator to behave badly. We shall see that this is indeed the case. Judge and Bock [21] were the first to provide a thorough discussion of pretest estimators in a regression context, but they did not notice the essential equivalence proved in section 2.

The *risk* of an estimator  $t(x; c)$  is defined as its MSE,

$$(4.4) \quad R(\theta, c) = \mathbf{E}_\theta(t(x; c) - \theta)^2,$$

where  $\mathbf{E}_\theta$  denotes expectation with respect to the  $N(\theta, 1)$  distribution.

**THEOREM 4.1.** *The risk of the pretest estimator  $t(x; c)$  is given by*

$$R(\theta, c) = 1 + (c + \theta) \varphi(c + \theta) + (c - \theta) \varphi(c - \theta) + (\theta^2 - 1) P(\theta, c),$$

where  $\varphi$  denotes the standard normal density and  $P(\theta, c) = \int_{-\theta-c}^{-\theta+c} \varphi(u) du$ .

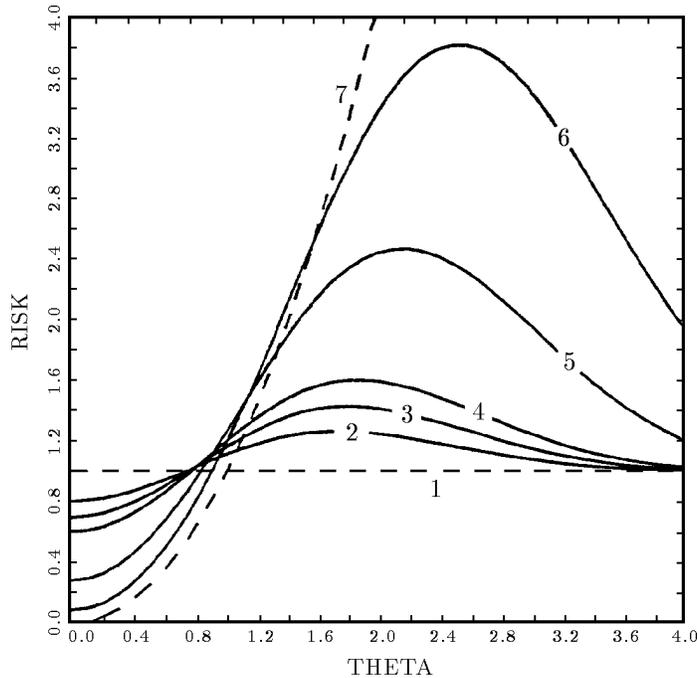


FIG. 2. Risk  $R(\theta, c)$  of the pretest estimator for seven values of  $c$ .

An alternative expression can be obtained from [5, Theorems A and B]; see also [21].

There appears to be some confusion about the admissibility of the pretest estimator. This confusion arises because in the class of pretest estimators no estimator dominates any other. But outside this class there are estimators which dominate the pretest estimator, because of its discontinuity; see [10]. (An estimator  $t$  dominates another estimator  $t'$  if  $MSE(t)$  is never larger, and at least at one point smaller, than  $MSE(t')$ . If no estimator dominates  $t$ , then  $t$  is *admissible*.) We have the following theorem.

**THEOREM 4.2.** *The estimator  $t(x; c)$ ,  $0 \leq c \leq \infty$ , is admissible if  $c = 0$  or  $c = \infty$  and is inadmissible otherwise. However, within the class of pretest estimators no estimator dominates any other.*

Some further properties of the pretest estimator are given in Theorem 4.3 below.

**THEOREM 4.3.** *The risk  $R(\theta, c)$  of the pretest estimator satisfies:*

- (a)  $R(\theta, c)$  is symmetric in  $\theta$ ;
- (b)  $R(\theta, 0) = 1$ ,  $R(\theta, \infty) = \theta^2$ ;
- (c)  $R(\theta, c)$  is bounded for every  $c < \infty$ ;
- (d) for every  $c < \infty$ ,  $R(\theta, c) \rightarrow 1$  as  $|\theta| \rightarrow \infty$ .

These properties are clarified in Figure 2, where  $R(\theta, c)$  is given as a function of  $\theta$  for seven different values of  $c$ , labeled 1–7. These values are, respectively, 0.0, 1.0, 1.2007, 1.3692, 1.96, 2.576,  $\infty$ . The figure confirms that no estimator in this class dominates any other (Theorem 4.2), and that the risk is bounded and converges to 1 as  $\theta \rightarrow \infty$  (unless  $c = \infty$ ).

Closer inspection of Figure 2 reveals a particularly damaging property of the pretest estimator: for  $\theta$  close to 0 we see, as expected, that the pretest estimator is

better than the “usual” estimator  $t(x; 0) = x$ , but worse than the “silly” estimator  $t(x; \infty) = 0$ . When  $\theta$  is large, the situation is reversed. This, again, is what we would expect. However, for moderate values of  $\theta$ , in particular around  $\theta = 1$ , we would like an improved estimator (such as the pretest estimator) to perform better (have lower risk) than both the “usual” and the “silly” estimator. Figure 2 shows that the opposite is the case! This result is not new — it even appears in some of the textbooks (see [23, p. 75]) — but its harmful consequences do not seem to have been fully appreciated. The pretest estimator is like the boat in section 3: it performs its worst where it should perform its best. In [33] we see that it is in fact possible to design a boat which performs well in all weather conditions, but this is not the boat designed by our hero. That is, there are estimators that either outperform both the “usual” and the “silly” estimators or are better than one and worse than the other. These estimators, unlike the pretest estimator, are therefore never worse than both the “usual” and the “silly” estimators.

The values for  $c$  in Figure 2 are chosen with care. We have  $c = 0$  and  $c = \infty$  as our extremes (in dotted lines). In between we have the important cases  $c = 1$ ,  $c = 1.96$ , and  $c = 2.576$ . These last two values correspond, of course, to the usual 5% and 1% levels of significance. There are two further graphs,  $c = 1.2007$  and  $c = 1.3692$ , corresponding to minimax regret solutions to be discussed later.

To obtain further insight we study the behavior of the risk function for  $c$  close to 0 and for  $\theta$  close to 0.

THEOREM 4.4. (a) For  $c$  close to 0,

$$R(\theta, c) = 1 + \frac{4}{3} \varphi(\theta) \left( \theta^2 - \frac{1}{2} \right) c^3 + O(c^5).$$

(b) For  $\theta$  close to 0,

$$R(\theta, c) = h_0(c) + h_1(c) \theta^2 + O(\theta^4),$$

where  $h_0(c) = 1 + 2c \varphi(c) - \int_{-c}^c \varphi(u) du$ ,  $h_1(c) = c(c^2 - 2) \varphi(c) + \int_{-c}^c \varphi(u) du$ .

(c)  $0 \leq h_0(c) \leq 1$ ,  $h_0'(c) \leq 0$ ,  $h_1(c) \geq 0$ ,  $h_1'(c) = 1$  for  $c = 1.6149$ , and  $c = \infty$  only.

There are four relations between  $\theta$  and  $c$  of particular interest. These are graphed in Figure 3. The first relation is  $R(\theta, c) = 1$ , giving all points where the risk of a pretest estimator equals the risk of the “usual” estimator ( $c = 0$ ). The relation can be solved explicitly for  $\theta$  in terms of  $c$ . That is, there exists a unique function  $\theta_0$ , defined on  $[0, \infty)$ , such that  $R(\theta_0(c), c) = 1$ . The function  $\theta_0$  is monotonically increasing and

$$(4.5) \quad \frac{1}{2} \sqrt{2} \leq \theta_0(c) \leq 1 \quad \text{for all } c \in (0, \infty],$$

using Theorem 4.4(a). The second relation is  $R(\theta, c) = \theta^2$ , which gives all points where the risk of a pretest estimator is equal to the risk of the “silly” estimator ( $c = \infty$ ). There exists a unique function  $\theta_\infty$ , defined on  $[0, \infty)$ , such that  $R(\theta_\infty(c), c) = \theta_\infty^2$ . The function  $\theta_\infty$  is monotonically increasing and  $\theta_\infty(c) \geq 1$  for all  $c \geq 0$ . Next we consider the relation  $R'_\theta(\theta, c) \equiv \partial R(\theta, c) / \partial \theta = 0$ , providing all points where  $R(\theta, c)$ , considered as a function of  $\theta$ , attains a maximum (or a minimum, but this does not occur). It follows from Theorem 4.4(a) that, for  $c$  close to 0,

$$(4.6) \quad R'_\theta(\theta, c) = -\frac{4}{3} \theta \left( \theta^2 - \frac{5}{2} \right) \varphi(\theta) c^3 + O(c^5).$$

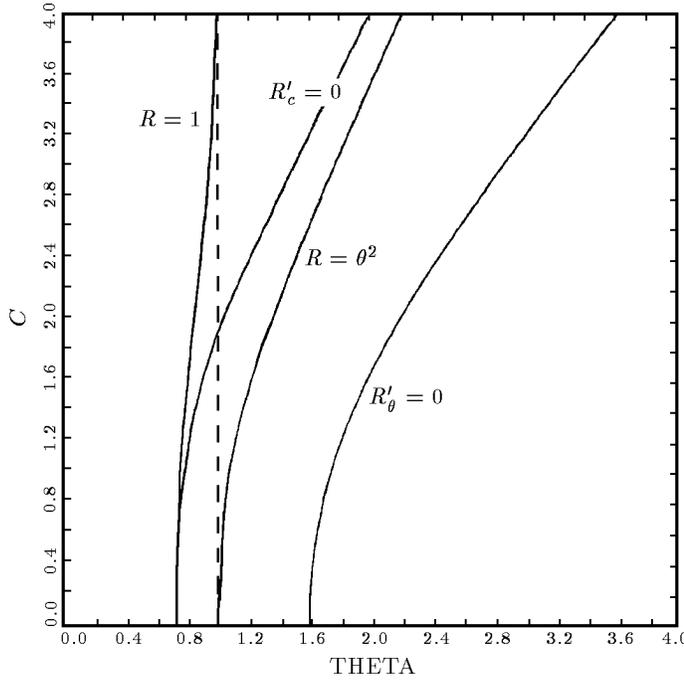


FIG. 3. Four relations between  $\theta$  and  $c$  for the pretest estimator.

Thus, there exists a unique function  $\theta_{\max}$  defined on  $(0, \infty)$  such that  $R(\theta_{\max}(c), c) = \sup_{\theta} R(\theta, c)$ . The function  $\theta_{\max}$  is monotonically increasing and  $\theta_{\max}(c) > \frac{1}{2}\sqrt{10}$  for  $c > 0$ . Furthermore,  $R(\theta_{\max}(c), c)$  is monotonically increasing in  $c$ . The fourth relation in Figure 3 is  $R'_c(\theta, c) \equiv \partial R(\theta, c)/\partial c = 0$ . At these points,  $R(\theta, c)$ , considered as a function of  $c$ , attains a maximum! Again applying Theorem 4.4(a) we have, for  $c$  close to 0,

$$(4.7) \quad R'_c(\theta, c) = 4\left(\theta^2 - \frac{1}{2}\right)\varphi(\theta)c^2 + O(c^4).$$

Thus, there exists a unique function  $c_{\max}$ , defined on  $[\frac{1}{2}\sqrt{2}, \infty)$ , such that

$$R(\theta, c_{\max}(\theta)) = \sup_c R(\theta, c).$$

The functions  $c_{\max}(\theta)$  and  $R(\theta, c_{\max}(\theta))$  are both monotonically increasing in  $\theta$ . With  $\theta_0$  and  $\theta_{\infty}$  as defined above we have, for every  $c > 0$ ,

$$(4.8) \quad \begin{aligned} \theta^2 < R(\theta, c) \leq 1 & \quad \text{if } |\theta| \leq \theta_0(c), \\ R(\theta, c) > 1 \geq \theta^2 & \quad \text{if } \theta_0(c) < |\theta| \leq 1, \\ R(\theta, c) \geq \theta^2 > 1 & \quad \text{if } 1 < |\theta| \leq \theta_{\infty}(c), \\ 1 < R(\theta, c) < \theta^2 & \quad \text{if } |\theta| > \theta_{\infty}(c). \end{aligned}$$

This shows again that there exists a range of  $\theta$ 's where the pretest estimator has higher risk than both the "usual" and the "silly" estimators.

An immediate consequence of (4.8) is Theorem 4.5 below.

THEOREM 4.5. *In the class of pretest estimators,  $\inf_c R(\theta, c) = \min(1, \theta^2)$ .*

In order to understand the importance of Theorem 4.5, let us consider once again the “usual” estimator ( $c = 0$ ,  $R(\theta, c) = 1$ ) and the “silly” estimator ( $c = \infty$ ,  $R(\theta, c) = \theta^2$ ). If we had to choose between these two we would choose the “usual” estimator if  $|\theta| > 1$  and the “silly” one if  $|\theta| \leq 1$ . (Of course,  $\theta$  is unknown.) If we now consider the whole class of pretest estimators, then Theorem 4.5 tells us that of all pretest estimators we would still choose the “usual” estimator  $x$  if  $|\theta| > 1$  and the “silly” estimator  $0$  if  $|\theta| \leq 1$ . Thus, in the context of pretest estimation, the only question which appears to be important is whether  $|\theta| \leq 1$  or  $|\theta| > 1$ . (We can go even one step further; see the discussion following Theorem 5.2.)

Given that  $|\theta| = 1$  appears as the natural pivot in the class of pretest estimators, we would like to choose  $c$  such that, at the very least, the estimator performs well (has low risk) around  $|\theta| = 1$ . One glance at Figure 3 tells us that the worst estimator in this respect is obtained around  $c = 1.9$  where the curve  $R'_c = 0$  crosses the line  $\theta = 1$ . (The exact value is  $c = 1.9150$ .) This corresponds to a significance level of 5.55%. Thus, ironically, the usual 5% estimator ( $c = 1.96$ ) is very close to being the worst choice of  $c$  in the sense that it gives the highest risk in the interval around  $\theta = 1$ !

**5. Regret.** If, in spite of the discussion in the previous section, we wish to use the traditional pretest estimator, how should we judge its performance? Obviously, we shall need the risk function  $R(\theta, c)$ . But since, in the class of pretest estimators, no estimator dominates the other (Theorem 4.2), some further criterion is needed. The minimax approach sometimes leads to unreasonable or trivial results [19]. Indeed, in our case, minimizing the maximum risk always leads to the “usual” estimator  $t(x; c) = x$ . On both theoretical and practical grounds [34], [9], [35], we shall adopt the minimax regret approach where we minimize the maximum regret instead of the maximum risk. The *regret* of an estimator  $t(x; c)$  is defined as

$$(5.1) \quad r(\theta, c) = R(\theta, c) - \inf_c R(\theta, c)$$

and an estimator  $t(x; c^*)$  is a *minimax regret* if

$$(5.2) \quad \sup_{\theta} r(\theta, c^*) = \inf_c \sup_{\theta} r(\theta, c).$$

We know from Theorem 4.5 that the minimum risk in the class of pretest estimators is  $\min(1, \theta^2)$ . Hence,

$$(5.3) \quad r(\theta, c) = R(\theta, c) - \min(1, \theta^2).$$

Sawa and Hiromatsu [35] first obtained the minimax regret solution, that is, they found  $c$  which minimizes  $\sup_{\theta} r(\theta, c)$ . See also [12], [8], and [11].

There is, however, another way to define the regret function. Instead of comparing the risk of the pretest estimator with the minimum risk in the class of pretest estimators, we can also compare it with the minimum risk in a much larger class of estimators. In this class, denoted  $\mathcal{L}^0$  and discussed further in [33], the minimum risk is  $\theta^2/(1 + \theta^2)$ . Hence, as an alternative to (5.3), we obtain

$$(5.4) \quad r_0(\theta, c) = R(\theta, c) - \frac{\theta^2}{1 + \theta^2}.$$

Of course,  $\theta^2/(1 + \theta^2)$  is smaller than  $\min(1, \theta^2)$ . We provide both minimax regret solutions. To prove that these solutions are unique, we notice that, for  $\theta$  close to 0,

both regret functions can be written as

$$(5.5) \quad r(\theta, c) = h_0(c) + (h_1(c) - 1)\theta^2 + O(\theta^4)$$

(see Theorem 4.4(b)), and hence both functions attain a local maximum at  $\theta = 0$  if and only if  $0 < c < 1.6149$  (Theorem 4.4(c)).

THEOREM 5.1 (minimax regret). *In the class of pretest estimators,*

(a) *the minimax regret estimator is obtained for  $c = 1.3692$  with minimax regret  $r^* = 0.5988$ ;*

(b) *the  $\mathcal{L}^0$ -minimax regret estimator is obtained for  $c = 1.2007$  with  $\mathcal{L}^0$ -minimax regret  $r_0^* = 0.6958$ .*

The risk functions associated with  $c = 1.3692$  and  $c = 1.2007$  are graphed in Figure 2, and are labeled 4 and 3.

TABLE 1  
Maximum  $\mathcal{L}^0$ -regret for seven pretest estimators.

$c$	0.0	1.0	1.2007	1.3692	1.96	2.576	$\infty$
Maximum regret	1.0	0.8013	0.6958*	0.8412	1.6450	2.9522	$\infty$

In Table 1 we present the maximum  $\mathcal{L}^0$ -regret of seven pretest estimators. In each case the regret function is  $r_0$ , given in (5.4). The maximum regret is minimized at  $c = 1.2007$  in accordance with Theorem 5.1(b).

As an alternative to finding the minimum (over  $c$ ) of the *maximum* (over  $\theta$ ) regret, we could try to obtain the minimum of the *average* regret. This assumes specification of a prior distribution  $\pi(\theta)$  of  $\theta$ . Since the difference between average regret and average risk does not depend on  $c$ , minimizing average regret is the same as minimizing average risk. In Bayesian parlance, average risk is the Bayes risk of an estimator  $t(x; c)$  with respect to  $\pi$ . In order to implement this approach we need to specify a prior distribution for  $\theta$ . Which prior should we choose? Toyoda and Wallace [39] considered this question and chose a diffuse prior for  $\theta$ . They then concluded that the “usual” estimator  $x$  is the minimum average regret estimator. Given our previous results, this is not surprising. A diffuse prior on  $\theta$  implies that  $\mathbf{P}\{|\theta| > 1\} = 1$  and we know that  $x$  has lowest risk when  $|\theta| > 1$ . In our view, a diffuse prior is not appropriate. For example, it puts equal weight on the intervals  $(100, 101)$  and  $(0, 1)$  and this is counterintuitive.

For a normal prior we have Theorem 5.2 below.

THEOREM 5.2 (minimum average risk). *Let  $\pi(\theta)$  be an  $N(\mu, \tau^2)$  prior density of  $\theta$ . Then*

(a) *the average risk (with respect to  $\pi$ ) of the pretest estimator is given by*

$$\mathbf{E}_\pi R(\theta, c) = 1 - (\alpha + \beta\tau^2)\varphi(\alpha) + (\beta + \alpha\tau^2)\varphi(\beta) + (\mu^2 + \tau^2 - 1) \int_\alpha^\beta \varphi(u) du,$$

where  $\alpha = -c - \mu/\sqrt{1 + \tau^2}$ ,  $\beta = c - \mu/\sqrt{1 + \tau^2}$ ;

(b) *the minimum average risk is*

$$\inf_c \mathbf{E}_\pi R(\theta, c) = \min(1, \mu^2 + \tau^2);$$

and

(c) *the minimum is obtained for*

$$c_\pi = \begin{cases} \infty & \text{if } \mu^2 + \tau^2 \leq 1, \\ 0 & \text{if } \mu^2 + \tau^2 > 1. \end{cases}$$

The surprising consequence of Theorem 5.2 is that, with a normal prior for  $\theta$ , the minimum average regret solution is the “usual” estimator  $x$  when  $\mathbf{E}\theta^2 > 1$  and the “silly” estimator 0 when  $\mathbf{E}\theta^2 < 1$ . This emphasizes once again that, for the traditional pretest estimator, the key issue is whether (the expected value of)  $\theta^2$  is larger or smaller than 1.

**6. Conclusions.** In this paper we have studied the traditional pretest estimator in some detail, because it is an estimator which is routinely used in econometrics and other areas of applied statistics. This does not mean that we believe it is a good estimator. On the contrary, we have emphasized its poor properties throughout.

Our framework is the estimation of the parameters of interest  $\beta$  in a linear regression model  $y = X\beta + \gamma z + u$ . The explanatory variables in  $X$  are regarded as belonging in the equation according to some theory and can be thought of as the minimum set of variables required to explain  $y$ . The explanatory variable  $z$ , however, is included only because the researcher believes this might lead to “better” estimators of  $\beta$ . The focus of our analysis is the estimation of one or several linear combinations of the parameters  $\beta$ .

We propose to estimate  $\beta$  as a weighted average of the unrestricted estimator  $b_u$  and the restricted estimator  $b_r$  (with  $\gamma = 0$ ), that is,  $b = \lambda b_u + (1 - \lambda)b_r$ , where  $\lambda$  is a function of the  $t$ -ratio of  $\gamma$ . This classical problem of estimating  $\beta$  in the presence of a nuisance parameter  $\gamma$  we have called *the regression problem*. In Theorem 2.2 (the equivalence theorem) we showed that the regression problem is equivalent to a fundamental statistical problem, which we have called the  $N(\theta, 1)$  *problem*: Given one observation  $x$  from an  $N(\theta, 1)$  distribution, what is the “best” estimator for  $\theta$ ? This seemingly trivial problem turns out to be far from trivial. It has been considered in detail by Magnus [33]. A wide range of possible estimators was considered, the only finally preferred having an interpretation as a Bayes estimator with prior density  $\frac{1}{2}c \exp(-c|\theta|)$ ,  $-\infty < \theta < \infty$ ,  $c = \log 2$ .

Because of the equivalence theorem, all properties of the pretest estimator in the regression problem follow from those of the “pretest” estimator,

$$\tilde{\theta} = \begin{cases} 0 & \text{if } |x| \leq c, \\ x & \text{if } |x| > c, \end{cases}$$

in the  $N(\theta, 1)$  problem. We have studied the pretest estimator in the context of risk and regret and we have seen that it is inadmissible (it is not even continuous) and has pathological behavior around  $|\theta| = 1$ . In particular, we conclude that the pretest estimator behaves badly exactly where we want it to behave well, that is, around  $|\theta| = 1$ ; see Theorems 2.1, 4.5, and 5.2.

The pretest estimator is like the boat of our hero in section 3. Compared with the two existing boats (the “usual” and the “silly” estimators) our hero’s boat (the pretest estimator) performs in-between when there is very little wind ( $|\theta|$  small) or under stormy conditions ( $|\theta|$  large), but it performs *worse* than either of the existing boats under normal conditions ( $|\theta|$  around 1).

**7. Proofs of theorems.**

*Proof of Theorems 2.1 and 2.2.* The theorems are special cases of results obtained in [32]. A direct proof of both theorems builds on the fact that the conditional distribution of  $(b_r, b_u)$  given  $\hat{\theta}$  is

$$\begin{pmatrix} b_r \\ b_u \end{pmatrix} \Big| \hat{\theta} \sim N \left[ \begin{pmatrix} \beta + \theta q \\ \beta - (\hat{\theta} - \theta)q \end{pmatrix}, \sigma^2 \begin{pmatrix} (X'X)^{-1} & (X'X)^{-1} \\ (X'X)^{-1} & (X'X)^{-1} \end{pmatrix} \right].$$

*Proof of Theorem 4.1.* With  $\lambda_c(x)$  defined in (4.3), we have

$$\begin{aligned} R(\theta, c) &= \int_{-\infty}^{\infty} (\lambda_c(x)x - \theta)^2 \varphi(x - \theta) dx \\ &= \theta^2 \int_{|x| \leq c} \varphi(x - \theta) dx + \int_{|x| > c} (x - \theta)^2 \varphi(x - \theta) dx \\ &= 1 + \theta^2 \int_S \varphi(u) du - \int_S u^2 \varphi(u) du \\ &= 1 + (\theta^2 - 1) P(\theta, c) - \int_S (u^2 - 1) \varphi(u) du, \end{aligned}$$

where  $S = \{u: -\theta - c < u < -\theta + c\}$ . Noting that  $\varphi'(u) = -u\varphi(u)$ ,  $\varphi''(u) = (u^2 - 1)\varphi(u)$ , we find

$$\int_S (u^2 - 1) \varphi(u) du = \int_S \varphi''(u) du = [\varphi'(u)]_S = -(c + \theta) \varphi(c + \theta) - (c - \theta) \varphi(c - \theta),$$

which concludes the proof.

*Proof of Theorem 4.2.* For  $c = 0$ , Blyth [4] showed that  $t(x; 0) \equiv x$  is admissible; see also [3, p. 548]. For  $c = \infty$ ,  $t(x; \infty) \equiv 0$  dominates every other estimator at  $\theta = 0$  and hence is admissible. For  $0 < c < \infty$ ,  $t(x; c)$  is discontinuous at  $x = \pm c$  and hence not admissible [3, pp. 542–544]. This proves the first part of the theorem. To prove the second part we follow Kempthorne [25]; see also [11]. Assume a prior distribution  $\pi(\theta)$  for  $\theta$ . In particular, let  $\pi(\theta)$  be  $N(0, 1)$ . Then we can show either directly or by using Theorem 5.2 that  $\mathbf{E}_\pi R(\theta, c) = 1$  for every  $c$ . This implies that there cannot exist two values  $c_1 \neq c_2$  such that  $R(\theta, c_1) \leq R(\theta, c_2)$  for all  $\theta$  with strict inequality for some  $\theta$ . Hence, no pretest estimator dominates any other. (See [3, pp. 253–254].)

*Proof of Theorem 4.3.* (a), (b), and (d) follow from Theorem 4.1. To prove (c) we note, from (4.2), that  $|t(x; c) - x| \leq c$  for all  $x$ . Hence,

$$\begin{aligned} R(\theta, c) &= \mathbf{E}(t(x; c) - \theta)^2 = \mathbf{E}\left((t(x; c) - x) + (x - \theta)\right)^2 \\ &\leq 2\mathbf{E}\left((t(x; c) - x)^2 + (x - \theta)^2\right) \leq 2(c^2 + 1). \end{aligned}$$

*Proof of Theorem 4.4.* For small  $c$  we have

$$\varphi(c + \theta) = \varphi(\theta) \left[ 1 - \theta c + \frac{1}{2} (\theta^2 - 1) c^2 - \frac{1}{6} \theta (\theta^2 - 3) c^3 \right] + O(c^4),$$

from which we obtain

$$\begin{aligned} \varphi(c + \theta) + \varphi(c - \theta) &= \varphi(\theta) [2 + (\theta^2 - 1) c^2] + O(c^4), \\ \varphi(c + \theta) - \varphi(c - \theta) &= -\varphi(\theta) \left[ 2\theta c + \frac{1}{3} \theta (\theta^2 - 3) c^3 \right] + O(c^5), \\ P(\theta, c) &= \varphi(\theta) \left[ 2c + \frac{1}{3} (\theta^2 - 1) c^3 \right] + O(c^5). \end{aligned}$$

Inserting these expansions in  $R(\theta, c)$ , given in Theorem 4.1, proves (a). The proof of (b) is similar. To prove (c) we notice that  $2c\varphi(c) \leq \int_{-c}^c \varphi(u) du \leq 1$  and hence  $0 \leq 2c\varphi(c) \leq h_0(c) \leq 1$  and  $h_1(c) \geq c^3\varphi(c) \geq 0$ . Also,  $h'_0(c) = -2c\varphi(c) \leq 0$  and  $h'_1(c) = c^2(5 - c^2)\varphi(c)$ . Since  $h_1(0) = 0$ ,  $h_1(\infty) = 1$ , it follows that there exists a unique  $c$ ,  $0 < c < \sqrt{5}$ , such that  $h(c) = 1$ . The value of this  $c$  can be approximated to any degree of accuracy.

*Proof of Theorem 4.5.* It follows from (4.8) that, for  $c \in (0, \infty)$ ,  $R(\theta, c) > \theta^2$  if  $|\theta| \leq 1$  and  $R(\theta, c) > 1$  if  $|\theta| > 1$ . The bounds are obtained for  $c = \infty$  and  $c = 0$ , respectively.

*Proof of Theorem 5.1.* A formal proof of the uniqueness of the minimax regret estimator can be found in [11]. For the case of  $\mathcal{L}^0$  the proof is similar, but tedious. Once we know that a unique solution exists, it can be found numerically to any desired degree of precision. (For readers who accept computer output as “proof,” the output shows unambiguously that there is a unique solution.)

*Proof of Theorem 5.2.* The proof of (a) is a straightforward exercise in integration. To prove (b) let us first consider the case  $\mu = 0$ . If  $\mu = 0$ , then  $\mathbf{E}_\pi R(\theta, c)$  reduces to  $\mathbf{E}_\pi R(\theta, c) = 1 + (\tau^2 - 1)\kappa(c)$ , where  $\kappa(c) = \int_{-\delta}^{\delta} \varphi(u) du - 2\delta\varphi(\delta)$ ,  $\delta = c/\sqrt{1 + \tau^2}$ .

Now,  $\kappa(c)$  is monotonically increasing on  $(0, \infty)$  with  $\kappa(0) = 0$ ,  $\kappa(\infty) = 1$ . Hence, if  $\tau^2 - 1 > 0$  the minimum is obtained by choosing  $\kappa(c) = 0$  (as small as possible), that is,  $c = 0$ . If  $\tau^2 - 1 < 0$  we must choose  $\kappa(c) = 1$  (as large as possible), that is,  $c = \infty$ . This proves (b) and (c) for the special case  $\mu = 0$ . The general case is rather tedious and is left to the reader.

**Acknowledgments.** Preliminary versions of this paper were presented at Tilburg University, Humboldt University, LSE, Yale, University of Amsterdam, University of Basel, the New Economic School of Moscow, and the Fifth World Meeting of ISBA, Istanbul 1997. I am grateful to the participants at these seminars. In particular, I thank Jim Berger, Bernd Droge, Ed Leamer, Peter Phillips, and Arnold Zellner for their useful comments, and Anurag Banerjee for his help in producing the graphs. My greatest thanks goes to Jim Durbin, without whose continuous encouragement and constructive criticism this paper would not have been written.

#### REFERENCES

- [1] T. A. BANCROFT, *On biases in estimation due to the use of preliminary tests of significance*, Ann. Math. Statist., 15 (1944), pp. 190–204.
- [2] T. A. BANCROFT, *Analysis and inference for incompletely specified models involving the use of preliminary tests of significance*, Biometrics, 20 (1964), pp. 427–442.
- [3] J. O. BERGER, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.
- [4] C. R. BLYTH, *On minimax statistical decision procedures and their admissibility*, Ann. Math. Statist., 22 (1951), pp. 22–42.
- [5] M. E. BOCK, *Minimax estimators of the mean of a multivariate normal distribution*, Ann. Statist., 3 (1975), pp. 209–218.
- [6] M. E. BOCK, G. G. JUDGE, AND T. A. YANCEY, *Some comments on estimation in regression after preliminary tests of significance*, J. Econometrics, 1 (1973), pp. 191–200.
- [7] M. E. BOCK, T. A. YANCEY, AND G. G. JUDGE, *The statistical consequences of preliminary test estimators in regression*, J. Amer. Statist. Assoc., 68 (1973), pp. 109–116.
- [8] R. J. BROOK, *On the use of a regret function to set significance points in prior tests of estimation*, J. Amer. Statist. Assoc., 71 (1976), pp. 126–131.
- [9] H. CHERNOFF AND L. E. MOSES, *Elementary Decision Theory*, Wiley, New York, 1959.

- [10] A. COHEN, *Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution*, Ann. Math. Statist., 36 (1965), pp. 78–87.
- [11] B. DROGE, *On finite sample properties of adaptive least squares regression estimates*, Statistics, 24 (1993), pp. 181–203.
- [12] R. W. FAREBROTHER, *Minimax regret significance points for a preliminary test in regression analysis: Comment*, Econometrica, 43 (1975), pp. 1005–1006.
- [13] M. S. FELDSTEIN, *Multicollinearity and the mean square error of alternative estimators*, Econometrica, 41 (1973), pp. 337–346.
- [14] M. FRIEDMAN, *Review of Jan Tinbergen. Statistical testing of business cycle theories, II: Business cycles in the United States of America*, Amer. Econ. Rev., 30 (1940), pp. 657–661.
- [15] J. A. GILES AND D. E. A. GILES, *Pre-test estimation and testing in econometrics: Recent developments*, J. Econ. Surveys, 7 (1993), pp. 145–197.
- [16] D. E. A. GILES AND A. C. RAYNER, *The mean squared errors of the maximum likelihood and natural-conjugate Bayes regression estimators*, J. Econometrics, 11 (1979), pp. 319–334.
- [17] J. GOODNIGHT AND T. D. WALLACE, *Operational techniques and tables for making weak MSE tests for restrictions in regressions*, Econometrica, 40 (1972), pp. 699–709.
- [18] T. HAAVELMO, *The probability approach in econometrics*, Econometrica, 12 (1944) (supplement), pp. 1–115.
- [19] J. L. HODGES AND E. L. LEHMANN, *Some problems in minimax point estimation*, Ann. Math. Statist., 21 (1950), pp. 182–197.
- [20] D. V. HUNTSBERGER, *A generalization of a preliminary testing procedure for pooling data*, Ann. Math. Statist., 26 (1955), pp. 734–743.
- [21] G. G. JUDGE AND M. E. BOCK, *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland, Amsterdam, 1978.
- [22] G. G. JUDGE AND M. E. BOCK, *Biased estimation*, in Handbook of Econometrics, Vol. I, Z. Griliches and M. D. Intriligator eds., North-Holland, Amsterdam, 1983, pp. 599–649.
- [23] G. G. JUDGE, W. E. GRIFFITHS, R. C. HILL, H. LUTKEPOHL, AND T. C. LEE, *The Theory and Practice of Econometrics*, Wiley, New York, 1985.
- [24] G. G. JUDGE AND T. A. YANCEY, *Improved Methods of Inference in Econometrics*, North-Holland, Amsterdam, 1986.
- [25] P. J. KEMPTHORNE, *Admissible variable-selection procedures when fitting regression models by least squares for prediction*, Biometrika, 71 (1984), pp. 593–597.
- [26] J. M. KEYNES, *Professor Tinbergen's method*, Econ. J., 49 (1939), pp. 558–568.
- [27] T. KOOPMANS, *Identification problems in economic model construction*, Econometrica, 17 (1949), pp. 125–144.
- [28] H. J. LARSON AND T. A. BANCROFT, *Biases in prediction by regression for certain incompletely specified models*, Biometrika, 50 (1963), pp. 391–402.
- [29] E. E. LEAMER, *Specification Searches*, Wiley, New York, 1978.
- [30] E. E. LEAMER, *Model choice and specification analysis*, in Handbook of Econometrics, vol. I, Z. Griliches and M. D. Intriligator, eds., North-Holland, Amsterdam, 1983, pp. 285–330.
- [31] E. E. LEAMER AND G. CHAMBERLAIN, *A Bayesian interpretation of pretesting*, J. Roy. Statist. Soc. Ser. B, 38 (1976), pp. 85–94.
- [32] J. R. MAGNUS AND J. DURBIN, *Estimation of regression coefficients of interest when other regression coefficients are of no interest*, Econometrica, 67 (1999), pp. 639–643.
- [33] J. R. MAGNUS, *Estimation of the mean of a univariate normal distribution with known variance*, Econometric Theory, submitted.
- [34] L. J. SAVAGE, *The theory of statistical decision*, J. Amer. Statist. Assoc., 46 (1951), pp. 55–67.
- [35] T. SAWA AND T. HIROMATSU, *Minimax regret significance points for a preliminary test in regression analysis*, Econometrica, 41 (1973), pp. 1093–1101.
- [36] S. L. SCLOVE, C. MORRIS, AND R. RADHAKRISHNAN, *Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution*, Ann. Math. Statist., 43 (1972), pp. 1481–1490.
- [37] J. TINBERGEN, *Statistical Testing of Business Cycle Theories*, Vols. 1, 2, League of Nations, Geneva, 1939.
- [38] C. TORO-VIZCARRONDO AND T. D. WALLACE, *A test of the mean square error criterion for restrictions in linear regression*, J. Amer. Statist. Assoc., 63 (1968), pp. 558–572.
- [39] T. TOYODA AND T. D. WALLACE, *Optimal critical values for pre-testing in regression*, Econometrica, 44 (1976), pp. 365–375.
- [40] T. D. WALLACE, *Efficiencies for stepwise regressions*, J. Amer. Statist. Assoc., 59 (1964), pp. 1179–1182.
- [41] T. D. WALLACE, *Weaker criteria and tests for linear restrictions in regression*, Econometrica, 40 (1972), pp. 689–698.

- [42] T. D. WALLACE AND V. G. ASHTAR, *Sequential methods in model construction*, Rev. Econ. Statist., 54 (1972), pp. 172–178.
- [43] T. D. WALLACE AND C. E. TORO-VIZCARRONDO, *Tables for the mean square error test for exact linear restrictions in regression*, J. Amer. Statist. Assoc., 64 (1969), pp. 1649–1663.
- [44] A. ZELLNER AND W. VANDAELE, *Bayes–Stein estimators for k-means, regression and simultaneous equation models*, in *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, S. E. Fienberg and A. Zellner, eds., North-Holland, Amsterdam, 1975, pp. 627–653.