P000372

pretesting

This article briefly discusses the meaning and dangers of pretesting in estimation procedures. It outlines the proof of the equivalence theorem, and compares the pretest estimator with three other estimators: the 'usual' estimator, the 'silly' estimator and the 'Laplace' estimator.

1 Model selection versus estimation

Suppose data are generated by a linear relationship

$$y = X\beta + \gamma z + u, \quad u \sim \mathrm{N}(0, \sigma^2 I_n), \tag{1}$$

where X is an $n \times k$ matrix of explanatory variables and z is an additional $n \times 1$ vector of explanatory variables. In our role as investigator we do not know this relationship. Our interest is in the effect of X on y, that is, we want to estimate β. Since we don't know that y is generated by (1), we formulate a model that will serve as a vehicle to estimate β. Let us assume that we know that the relationship is linear, that X is certainly in the model, and that z is perhaps in the model. For simplicity we assume also that σ^2 is known. Thus our 'model space' consists of only two models: the unrestricted model (where $\gamma \neq 0$) and the restricted model (where $\gamma = 0$).

Our interest could be in finding the 'true' model, in which case we are concerned with *model selection*. In that case we should select the unrestricted model, however small γ turns out to be. Our interest, however, is in the *estimation* of β and the model is not of interest *per se* – it is only a means towards our goal. Even if we knew that γ is nonzero, this would not necessarily mean that we should include z in our regression. This is because, if γ is close to zero, a small bias in the estimates of β will result if we use the restricted model, but their variances may increase substantially, and hence the mean squared error will also increase substantially. (Recall that the bias depends on the value of γ but the variance does not.) So even if we know the truth, it is typically wise to simplify for the purposes of estimation.

2 What is pretesting?

The ordinary least squares (OLS) estimator for β in the restricted model is of course

$$b_r = (X'X)^{-1}X'y. \tag{2}$$

If we define

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

then we can write the OLS estimators for β and γ in the unrestricted model as

$$b_u = b_r - \hat{\theta}q, \quad \hat{\gamma} = \frac{z'My}{z'Mz}, \tag{3}$$

where

$$\hat{\theta} = \frac{\hat{\gamma}}{\sigma/\sqrt{z'Mz}} = \frac{z'My}{\sigma\sqrt{z'Mz}} \sim \mathrm{N}(\theta, 1) \tag{4}$$

denotes the t-ratio, which is normally distributed in this case because σ^2 is assumed known. We call θ the *theoretical* t-ratio.

Since we don't know which of the two models we should use in order to estimate β, the typical econometric practice is to perform a preliminary test (pretest) on γ, and to include z in our regression if the t-ratio $\hat{\theta}$ is 'large' and exclude it if $\hat{\theta}$ is 'small'. This leads to the so-called *pretest estimator*

$$b = \begin{cases} b_r & \text{if} \quad |\hat{\theta}| \le c, \\ b_u & \text{if} \quad |\hat{\theta}| > c, \end{cases} \tag{5}$$

where c is some positive number such as 1.96. We can also write (5) as

$$b = \lambda b_u + (1-\lambda) b_r, \quad \lambda = \begin{cases} 0 & \text{if} \quad |\hat{\theta}| \le c, \\ 1 & \text{if} \quad |\hat{\theta}| > c, \end{cases} \tag{6}$$

which emphasizes that the pretest estimator is a weighted average of the estimators in the available models. The weights, however, are random variables because they depend on $\hat{\theta}$. The pretest estimator is therefore a complicated nonlinear estimator.

The problem with pretesting is not so much that people do it, but that they ignore the consequences. In typical econometric practice, model selection takes place using t-ratios and other diagnostics, after which a single model is selected (stage 1). Then estimates and standard errors are obtained in the selected model (stage 2), and these are reported. It is then tacitly assumed that the reported estimates are unbiased and that their standard errors are given by the usual OLS formulae. This assumption, however, is incorrect. The estimates are biased and their standard errors are not given by the usual OLS formulae. This is the pretest problem.

3 The equivalence theorem

Things are made simpler by the equivalence theorem, originally proved by Magnus and Durbin (1999), and improved and extended by Danilov and Magnus (2004a).

Theorem 1 (Equivalence theorem): Let $b = \lambda b_u + (1-\lambda) b_r$, where $0 \le \lambda \le 1$ and $\lambda = \lambda(My)$.

Then, letting $\tilde{\theta} = \lambda\hat{\theta}$, we have

$$E(b) = \beta - E(\tilde{\theta} - \theta)q, \quad var(b) = \sigma^2 (X'X)^{-1} + var(\tilde{\theta})qq'$$

and hence

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

Proof We know from (3) that $b_u = b_r - \hat{\theta}q$, so that

$$b = \lambda b_u + (1-\lambda) b_r = b_r - \lambda\hat{\theta}q = b_r - \tilde{\theta}q.$$

The crucial ingredient is that b_r and My are independent, so that

$$E(b_r|My) = E(b_r), \quad var(b_r|My) = var(b_r).$$

Also, since both λ (by assumption) and $\hat{\theta}$ as given in (4) depend only on My, we see that $\tilde{\theta} = \lambda\hat{\theta}$ depends only on My. Hence,

$$E(b|My) = E(b_r) - E(\tilde{\theta}|My)q = \beta + \theta q - \tilde{\theta}q = \beta - (\tilde{\theta} - \theta)q$$

and

$$var(b|My) = var(b_r|My) = var(b_r) = \sigma^2(X'X)^{-1}.$$

Now using the well-known relationships between conditional and unconditional moments, we obtain

$$E(b) = E(E(b|My)) = \beta - E(\tilde{\theta} - \theta)q,$$

and

$$\begin{aligned} var(b) &= E(var(b|My)) + var(E(b|My)) \\ &= \sigma^2(X'X)^{-1} + var(\tilde{\theta})qq', \end{aligned}$$

and hence

$$\begin{aligned} MSE(b) &= var(b) + E(b-\beta)(b-\beta)' \\ &= \sigma^2(X'X)^{-1} + var(\tilde{\theta})qq' + (E(\tilde{\theta} - \theta))^2 qq' \\ &= \sigma^2(X'X)^{-1} + MSE(\tilde{\theta})qq'. \end{aligned}$$

This completes the proof. ||

The equivalence theorem is important because it tells us that if we have a 'good' estimator for θ, say $\tilde{\theta}$, then this defines $\lambda = \tilde{\theta}/\hat{\theta}$ and *the same* λ will provide a good estimator for β, namely $b = \lambda b_u + (1-\lambda)b_r$. The pretest estimator chooses

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

which is not a good choice as we shall see.

4 Moments of the pretest estimator

In the previous section we have seen that the pretest estimator is, in essence, of the form

$$t(x) = \begin{cases} 0 & \text{if} \quad |x| \le c, \\ x & \text{if} \quad |x| > c, \end{cases} \tag{7}$$

where $x \sim \mathrm{N}(\theta, 1)$. When studying this estimator, we confront it with three other estimators: the 'usual' estimator $t(x) = x$, the 'silly' estimator $t(x) = 0$, and the 'Laplace' estimator introduced in Magnus (2002). The four estimators are graphed in Figure 1 for $|x| < 4$.

It is clear that the pretest estimator is discontinuous, hence inadmissible. But this is only one of its uncomfortable properties.

Theorem 2 (Moments of pretest estimator): Let $x \sim \mathrm{N}(\theta, 1)$ and let $t(x)$ be the pretest estimator defined in (7). Then,

$$E(t-\theta) = \varphi(c-\theta) - \varphi(c+\theta) - \theta P$$

and

$$\begin{aligned} E(t-\theta)^2 = 1 &+ (c+\theta)\varphi(c+\theta) \\ &+ (c-\theta)\varphi(c-\theta) + (\theta^2 - 1)P, \end{aligned}$$

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

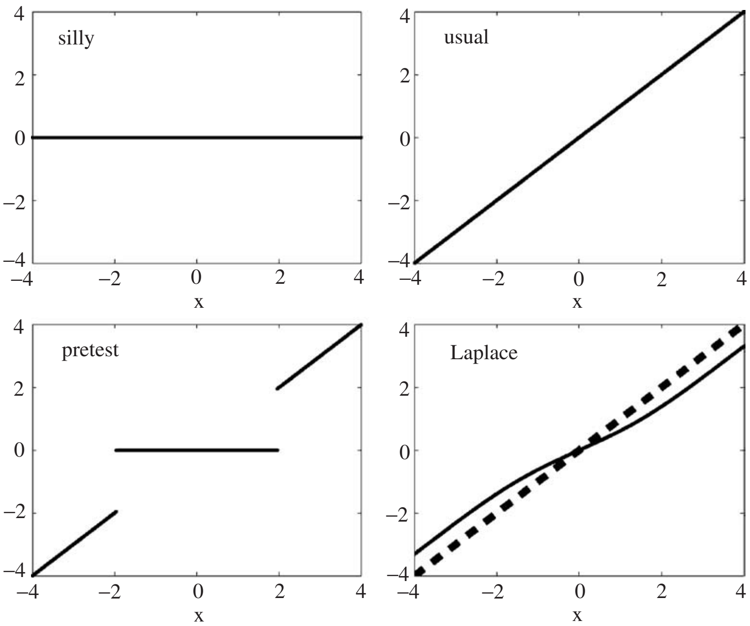

Figure 1 Four estimators $t(x)$ of θ.

Proof Letting $S = \{u : -\theta - c < u < -\theta + c\}$, we have

$$\begin{aligned} E(t(x)) &= \int_{-\infty}^{\infty} t(x)\varphi(x-\theta)dx = \int_{|x|>c} x\varphi(x-\theta)dx \\ &= \theta - \int_{|x|<c} x\varphi(x-\theta)dx = \theta - \int_S (u+\theta)\varphi(u)du \\ &= \theta - \int_S u\varphi(u)du - \theta \int_S \varphi(u)du \\ &= \theta + [\varphi(u)]_S - \theta P = \theta + \varphi(-\theta + c) - \varphi(-\theta - c) - \theta P, \end{aligned}$$

using the fact that $\varphi'(u) = -u\varphi(u)$. Similarly, using the fact that $\varphi''(u) = (u^2 - 1)\varphi(u)$, we obtain the second result. ||

The bias, standard error and root mean squared error of the pretest estimator are graphed in Figure 2.

We see that the bias is relatively small compared with the standard error. Since $bias(-\theta) = -bias(\theta)$ (so that θ and $bias(\theta)$ have opposite signs), and since we know from Theorem 1 that $bias(b_i) = -bias(\tilde{\theta})q_i$, we can determine the direction of the pretest bias.

Theorem 3 (Sign of pretest bias): Let $w := (X'X)^{-1}X'z$ with components w_i $(i = 1, \ldots, k)$. Then the pretest bias of b_i is positive (that is, $E(b_i) > \beta_i$) if and only if $\gamma_i w_i > 0$. As a consequence we can estimate the sign of the pretest bias of b_i by $sign(w_i\hat{\gamma}_i)$.

For purposes of exposition we have concentrated on the simplest case, but considerable generalization is possible to more than one additional z-variable, to unknown σ^2, and to general variance matrix.

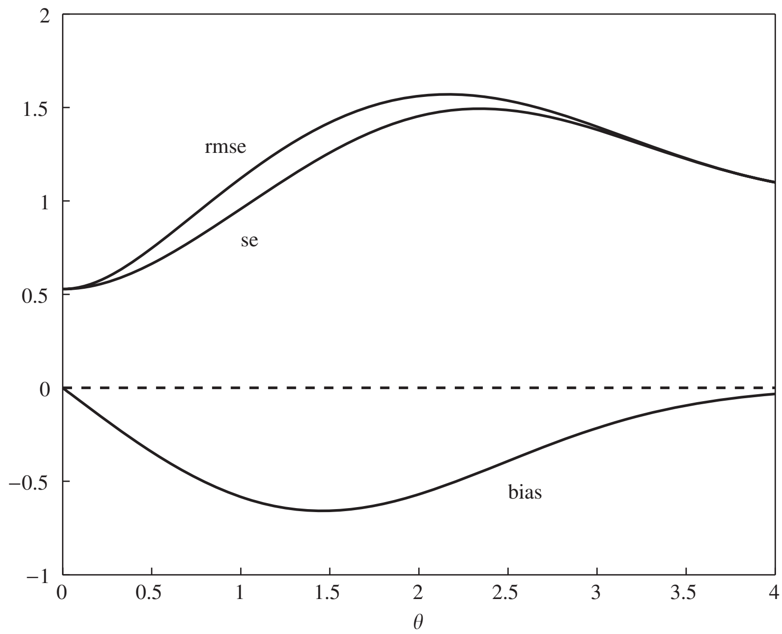

Figure 2 Moments of the pretest estimator.

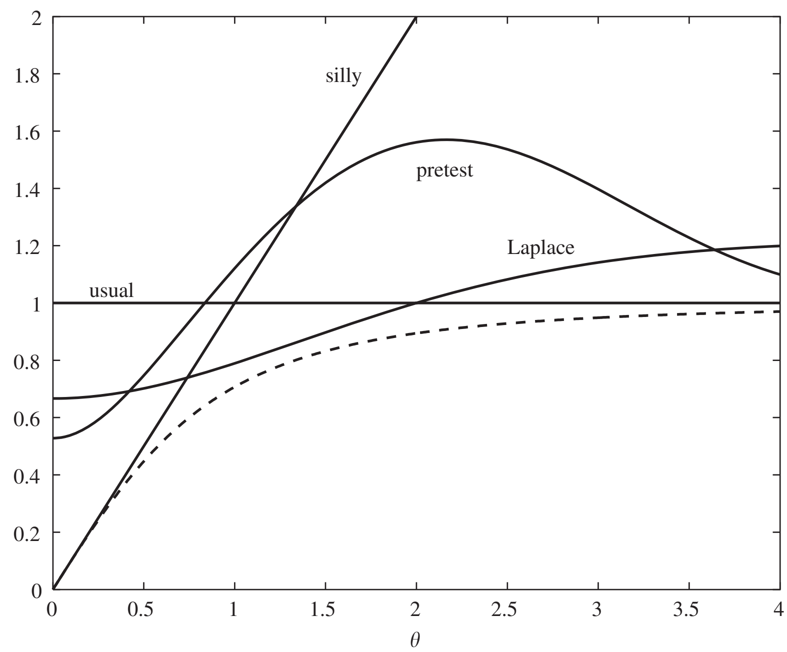

Figure 3 Root mean squared error of the four estimators.

5 Alternatives

We now compare the pretest estimator with the four estimators in Figure 1. We graph the root mean squared error (RMSE) of each of the four estimators in Figure 3.

The ‘usual’ estimator is unbiased and has variance one, independent of the value of θ. The ‘silly’ estimator is obviously better when θ is close to zero, the two estimators have the same RMSE when $\theta = 1$, corresponding to the fact that

$$MSE(b_r) - MSE(b_u) = (\theta^2 - 1)qq',$$

but the RMSE of the ‘silly’ estimator is unbounded. The pretest estimator lies in-between the silly and the usual estimator, except in the important interval around $\theta = 1$ where the pretest estimator is *worse* rather than better than either of the two naive alternatives. This is a most unwelcome property of the pretest estimator, and it has given rise to thought about alternatives. An attractive alternative is the so-called Laplace estimator, which has a Bayesian and a non-Bayesian interpretation, is admissible, is based on a ‘neutral’ prior, and has good properties around $\theta = 1$. The dotted line $|\theta|/\sqrt{1+\theta^2}$ denotes the theoretical lower bound of the root mean squared error.

6 History

The implications of model selection on the estimators of the parameters of interest were already being discussed following Tinbergen’s (1939) study for the League of Nations. Both Keynes (1939) and Friedman (1940), in their respective critiques on Tinbergen, focused on the method of model selection when the estimation procedure repeatedly uses the same data to discriminate between plausible competing theories. The same point was made in Haavelmo (1944, Section 17). Koopmans (1949) suggested that a completely new theory of inference was required to solve the dilemmas implied by the model selection problem.

Early work on the pretest estimator includes Bancroft (1944, 1964), Huntsberger (1955), Larson and Bancroft (1963), Cohen (1965), Wallace and Ashtar (1972), Sclove, Morris and Radhakrishnan (1972), Bock, Yancey and Judge (1973), and Bock, Judge and Yancey (1973). The harm of ignoring the effects of pretesting was analysed by Danilov and Magnus (2004a, 2004b). Important surveys are provided by Judge and Bock (1978, 1983), Judge and Yancey (1986), Giles and Giles (1993), and Magnus (1999).

Acknowledgements

I am grateful to Chris Müris for preparing the figures.

Jan R. Magnus

See also

< xref = M000379 > model selection;
< xref = R000254 > robust estimators in econometrics;
< xref = S000469 > semiparametric estimation;
< xref = T000203 > testing.

Bibliography

Bancroft, T.A. 1944. On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics* 15, 190–204.

Bancroft, T.A. 1964. Analysis and inference for incompletely specified models involving the use of preliminary tests of significance. *Biometrics* 20, 427–42.

Bock, M.E., Judge, G.G. and Yancey, T.A. 1973. Some comments on estimation in regression after preliminary tests of significance. *Journal of Econometrics* 1, 191–200.

Bock, M.E., Yancey, T.A. and Judge, G.G. 1973. The statistical consequences of preliminary test estimators in regression. *Journal of the American Statistical Association* 68, 109–16.

Cohen, A. 1965. Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution. *Annals of Mathematical Statistics* 36, 78–87.

Danilov, D. and Magnus, J.R. 2004a. On the harm that ignoring pretesting can cause. *Journal of Econometrics* 122, 27–46.

Danilov, D. and Magnus, J.R. 2004b. Forecast accuracy after pretesting with an application to the stock market. *Journal of Forecasting* 23, 251–74.

Friedman, M. 1940. Review of Jan Tinbergen, Statistical Testing of Business Cycle Theories, II: Business Cycles in the United States of America. *American Economic Review* 30, 657–61.

Giles, J.A. and Giles, D.E.A. 1993. Pre-test estimation and testing in econometrics: recent developments. *Journal of Economic Surveys* 7, 145–97.

Haavelmo, T. 1944. The probability approach in econometrics. *Econometrica* 12(supplement), 1–115.

Huntsberger, D.V. 1955. A generalization of a preliminary testing procedure for pooling data. *Annals of Mathematical Statistics* 26, 734–43.

Judge, G.G. and Bock, M.E. 1978. *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North-Holland.

Judge, G.G. and Bock, M.E. 1983. Biased estimation. In *Handbook of Econometrics*, Vol. 1, ed. Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland, Chapter 10.

Judge, G.G. and Yancey, T.A. 1986. *Improved Methods of Inference in Econometrics*. Amsterdam: North-Holland.

Keynes, J.M. 1939. Professor Tinbergen's method. *Economic Journal* 49, 558–68.

Koopmans, T. 1949. Identification problems in economic model construction. *Econometrica* 17, 125–44.

Larson, H.J. and Bancroft, T.A. 1963. Biases in prediction by regression for certain incompletely specified models. *Biometrika* 50, 391–402.

Magnus, J.R. 1999. The traditional pretest estimator. *Theory of Probability and its Applications* 44, 293–308.

Magnus, J.R. 2002. Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal* 5, 225–36.

Magnus, J.R. and Durbin, J. 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* 67, 639–43.

Sclove, S.L., Morris, C. and Radhakrishnan, R. 1972. Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Annals of Mathematical Statistics* 43, 1481–90.

Tinbergen, J. 1939. *Statistical Testing of Business Cycle Theories*, 2 vols. Geneva: League of Nations.

Wallace, T.D. and Ashtar, V.G. 1972. Sequential methods in model construction. *Review of Economics and Statistics* 54, 172–8.