Econometrica, Vol. 67, No. 3 (May, 1999), 639–643

ESTIMATION OF REGRESSION COEFFICIENTS OF INTEREST WHEN OTHER REGRESSION COEFFICIENTS ARE OF NO INTEREST

BY JAN R. MAGNUS AND J. DURBIN[1]

1. INTRODUCTION

WE ARE CONCERNED WITH THE ESTIMATION of (linear combinations of the elements of) β in the linear regression model $y = X\beta + Z\gamma + u$, where $y(n \times 1)$ is the vector of observations, $X(n \times k)$ and $Z(n \times m)$ are matrices of nonrandom regressors, $u(n \times 1)$ is a random vector of unobservable disturbances, and $\beta(k \times 1)$ and $\gamma(m \times 1)$ are unknown nonrandom parameters. We assume that $k \geq 1$, $m \geq 1$, $n - k - m \geq 1$, that the design matrix $(X : Z)$ has full column-rank $k + m$, and that the disturbances $u_1, u_2, \ldots, u_n$ are i.i.d. $N(0, \sigma^2)$. The reason for distinguishing between X and Z is that X contains the explanatory variables that are required to be in the model on theoretical or other grounds, while Z contains additional explanatory variables about which there is doubt as to whether they should be in the model or not. We are interested in estimating β (or specified linear combinations of its elements), while γ 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 β. We assess the relative performance of estimators by the mean squared error (MSE) criterion.

We introduce the matrices

$$M = I_n - X(X'X)^{-1}X' \quad \text{and} \quad Q = (X'X)^{-1}X'Z(Z'MZ)^{-1/2},$$

and an unknown parameter vector $\theta = (Z'MZ)^{1/2}\gamma$, which will play a central role in our analysis. The least squares (LS) estimators of β and γ are $b_u = b_r - Q\hat{\theta}$ and $\hat{\gamma} = (Z'MZ)^{-1}Z'My$, where $b_r = (X'X)^{-1}X'y$ and $\hat{\theta} = (Z'MZ)^{1/2}\hat{\gamma}$. The subscripts '$u$' and '$r$' stand for 'unrestricted' and 'restricted' (with $\gamma = 0$), respectively. In Section 2 we compare the mean squared errors of b_r and b_u. Theorem 1 gives precise conditions under which $\text{MSE}(b_r) - \text{MSE}(b_u)$ is positive semidefinite and negative semidefinite respectively. In the traditional approach to estimating β a choice is made between b_r and b_u depending on the outcome of a test of the hypothesis $\theta = 0$ (pretest estimation). A smoother and more appealing procedure is to estimate β as a weighted average of b_u and b_r, that is, $b = \lambda b_u + (1 - \lambda)b_r$, where $\lambda = \lambda(\hat{\theta}, s^2)$ is a scalar function of $\hat{\theta}$ (the LS estimator of θ) and s^2 (the LS estimator of σ^2). Section 3 studies the mean squared error of b. Let $\tilde{\theta} = \lambda(\hat{\theta}, s^2)\hat{\theta}$. Remarkably, the MSE of b, regarded as a function of $\lambda(\hat{\theta}, s^2)$, depends only on the MSE of $\tilde{\theta}$. Now $\hat{\theta}$ is a $N(\theta, \sigma^2 I_m)$ vector. This means that the task of finding a minimum MSE estimator of the coefficient vector β is equivalent to a problem that has nothing whatever to do with the structure of the regression model, namely the estimation of θ given a single observation from a $N(\theta, \sigma^2 I_m)$ distribution.

[1] 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. We are grateful to the participants at these seminars. In particular, we thank Ed Leamer, Peter Phillips, Arnold Zellner, a co-editor and two referees of *Econometrica* for their useful comments.

We found this result very surprising. The equivalence is stated explicitly in Theorem 2. The significance of the result is that if an acceptable solution can be found for the $N(\theta, \sigma^2 I_m)$ problem, it would apply to all linear regression models with the same value of m irrespective of the values of the regressors in the models.[2]

2. MSE COMPARISON OF b_r AND b_u

Since the joint distribution of $(b_r, b_u, \hat{\theta})$ is given by

$$\begin{pmatrix} b_r \\ b_u \\ \hat{\theta} \end{pmatrix} \sim N\left[\begin{pmatrix} \beta + Q\theta \\ \beta \\ \theta \end{pmatrix}, \sigma^2 \begin{pmatrix} (X'X)^{-1} & (X'X)^{-1} & 0 \\ (X'X)^{-1} & (X'X)^{-1} + QQ' & -Q \\ 0 & -Q' & I_m \end{pmatrix}\right], \tag{1}$$

we obtain the following result.

THEOREM 1: *We have*

$$MSE(b_r) - MSE(b_u) = Q\theta\theta'Q' - \sigma^2 QQ',$$

and hence

$$MSE(b_r) \leq MSE(b_u) \quad \text{iff} \quad \frac{\theta'Q'(QQ')^{-}Q\theta}{\sigma^2} \leq 1,$$

$$MSE(b_u) \leq MSE(b_r) \quad \text{iff} \quad r(Q) = 1 \quad \text{and} \quad \frac{\theta'Q'(QQ')^{-}Q\theta}{\sigma^2} \geq 1.$$

In particular, if $r(X'Z) = r(Z)$ so that $r(Q) = m$, then

$$MSE(b_r) \leq MSE(b_u) \quad \text{iff} \quad \theta'\theta/\sigma^2 \leq 1,$$

$$MSE(b_u) \leq MSE(b_r) \quad \text{iff} \quad m = 1 \quad \text{and} \quad \theta^2/\sigma^2 \geq 1.$$[3]

Note: For any two positive semidefinite matrices B and C, the notation $B \leq C$ means that $C - B$ is positive semidefinite. The notation A^{-} denotes a generalized inverse of A, that is, any matrix A satisfying $AA^{-}A = A$.

PROOF: The first statement is a direct consequence of (1). The next two statements follow from Theorem A1 in the Appendix, using the fact that $Q\theta$ lies in the column space of QQ'. Finally, if Q has full column-rank m, then $Q'(QQ')^{-}Q = I_m$ and the final two statements follow. *Q.E.D.*

[2] For the special case $m = 1$ with σ^2 known the problem reduces to the following: given a single observation from a $N(\theta, 1)$ distribution, find the best estimator of θ. This seemingly trivial problem is far from trivial. It has been considered in detail by Magnus and Durbin (1996) and by Magnus (1999b). A wide range of possible estimators was considered, the one finally preferred having an interpretation as a Bayes estimator with prior density $\frac{1}{2}c\exp(-c|\theta|)$, $-\infty < \theta < \infty$, $c = \log 2$.

[3] The special case $m = 1$ is "well-known" in the sense that, in some form, it has been around for a long time. The earliest reference is possibly exercise 12 in Wold (1953, p. 246), which he attributes to J. Durbin. See also Leamer (1978, p. 132).

Since σ^2 is unknown, let s^2 be its usual least squares estimator in the unrestricted model, $s^2 = (y - Xb_u - Z\hat{\gamma})'(y - Xb_u - Z\hat{\gamma})/(n-k-m)$, and notice that s^2 is independent of $(b_r, b_u, \hat{\theta})$. If we restrict ourselves to two possible estimators of β, b_r and b_u, then the traditional approach to the estimation of β is to base the choice between b_r and b_u on the statistic $F = \hat{\theta}'\hat{\theta}/(ms^2)$, which under the null hypothesis that $\gamma = 0$ follows an F distribution with m and $n-k-m$ degrees of freedom. The estimator b of β can then be written as $b = b_r$ if $F \le c$ and $b = b_u$ if $F > c$ for some $c \ge 0$. This estimator is called the (traditional) pretest estimator.[4] The F test asks the question: Is it true that $\theta = 0$? But this is the wrong question in this context. The right question is: What is the best available estimator of β?

Let us rewrite the pretest estimator as

$$b = \lambda b_u + (1-\lambda) b_r, \tag{2}$$

where $\lambda = 0$ if $F \le c$ and $\lambda = 1$ if $F > c$. This formulation shows b as a weighted average of b_u and b_r. A more general and more appealing class of estimators of β can be obtained by letting $\lambda = \lambda(\hat{\theta}, s^2)$ be a scalar function of $\hat{\theta}$ and s^2, such that $0 \le \lambda \le 1$. Any estimator of β of the form (2), where $\lambda = \lambda(\hat{\theta}, s^2)$ and $0 \le \lambda \le 1$, 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. In the next section we derive the mean squared error of the WALS estimator and show that it depends crucially on the mean squared error of $\lambda\hat{\theta}$, regarded as an estimator of θ.

3. THE EQUIVALENCE THEOREM

Using (1) and standard results on the multivariate normal distribution (Rao (1973, p. 522)), we obtain the conditional distribution of (b_r, b_u) given $\hat{\theta}$ as

$$\begin{pmatrix} b_r \\ b_u \end{pmatrix} \Bigg| \, \hat{\theta} \sim N\left[\begin{pmatrix} \beta + Q\theta \\ \beta - Q(\hat{\theta} - \theta) \end{pmatrix}, \sigma^2 \begin{pmatrix} (X'X)^{-1} & (X'X)^{-1} \\ (X'X)^{-1} & (X'X)^{-1} \end{pmatrix}\right]. \tag{3}$$

Since s^2 is independent of (b_r, b_u) and recalling that $b = \lambda b_u + (1-\lambda) b_r$ with $\lambda = \lambda(\hat{\theta}, s^2)$, we next find the conditional distribution of b given $\hat{\theta}$ and s^2 as

$$b \mid \hat{\theta}, s^2 \sim N(\beta - Q(\tilde{\theta} - \theta), \sigma^2 (X'X)^{-1}), \tag{4}$$

where $\tilde{\theta} = \lambda\hat{\theta}$. The unconditional mean and variance of the WALS estimator b are immediate consequences of (4):

$$Eb = \beta - QE(\tilde{\theta} - \theta), \quad \operatorname{var}(b) = \sigma^2 (X'X)^{-1} + Q \operatorname{var}(\tilde{\theta}) Q'. \tag{5}$$

From (5) we then obtain our main result.

THEOREM 2 (The Equivalence Theorem): *Let $b = \lambda b_u + (1-\lambda) b_r$, where $\lambda = \lambda(\hat{\theta}, s^2)$. Then,*

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

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

[4] See the surveys by Judge and Bock (1978, 1983) and Magnus (1999a).

The importance of Theorem 2 is that if we can find a λ-function such that $\lambda\hat{\theta}$ is an optimal estimator of θ, then *the same* λ-function will provide an optimal WALS estimator of β. The problem of estimating β in a regression context is thus reduced to estimating θ from a single vector observation $\hat{\theta} \sim N(\theta, \sigma^2 I_m)$. Although we have emphasized that we are interested in the parameters β and not in θ (that is, γ), which is a nuisance parameter, we now see, ironically and surprisingly, that optimal WALS estimation of β depends precisely on finding the optimal estimator of θ.

CentER for Economic Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; magnus@kub.nl,

and

London School of Economics and Political Science, Department of Statistics, Houghton Street, London WC2A, United Kingdom.

Manuscript received July, 1996; final revision received May, 1998.

APPENDIX: PROPERTIES OF THE MATRIX $A - aa'$, WHERE A IS POSITIVE SEMIDEFINITE

Let A be a positive semidefinite $n \times n$ matrix and a an $n \times 1$ vector. In this Appendix we obtain the rank of the matrix $A - aa'$ and necessary and sufficient conditions for $A - aa'$ to be positive semidefinite and negative semidefinite. Part (ii) of Theorem A1 is also given in Toutenburg (1992, Theorem 5.6), but without proof. Toutenburg attributes (ii) to Baksalary and Kala (1983).

THEOREM A1: *Let A be a positive semidefinite $n \times n$ matrix of rank $m \geq 1$, and let a be an $n \times 1$ vector. Let $B = A - aa'$. Then,*

(i) *the rank of B is*

$$r(B) = \begin{cases} m-1, & \text{if } r(A:a) = m \text{ and } a'A^- a = 1, \\ m, & \text{if } r(A:a) = m \text{ and } a'A^- a \neq 1, \\ m+1, & \text{if } r(A:a) = m+1; \end{cases}$$

(ii) *B is positive semidefinite if and only if $r(A:a) = m$ and $a'A^- a \leq 1$;*

(iii) *B is negative semidefinite if and only if $r(A:a) = 1$ and $a'A^- a \geq 1$.*

PROOF: First assume that A is diagonal, $A = \text{diag}(1, 1, \ldots, 1, 0, \ldots, 0)$, with m ones and $n - m$ zeros on the diagonal. Let $a' = (a_1', a_2')$ and $x' = (x_1', x_2')$ be partitioned accordingly. Then every x with $x_1 = 0$ and $a_2' x_2 = 0$ satisfies $Bx = 0$. Similarly, every x with $x_2 = 0$ and $a_1' x_1 = 0$ satisfies $Bx = x$. Hence B possesses (at least) $n - m - 1$ eigenvalues 0 and $m - 1$ eigenvalues 1. Let λ_1 and λ_2 denote the remaining two eigenvalues. From the equations

$$\text{tr}\, B = \lambda_1 + \lambda_2 + m - 1, \qquad \text{tr}\, B^2 = \lambda_1^2 + \lambda_2^2 + m - 1$$

we find

$$\lambda_{1,2} = \frac{1 - a_1'a_1 - a_2'a_2 \pm \sqrt{(1 - a_1'a_1 - a_2'a_2)^2 + 4a_2'a_2}}{2}$$

and hence

$$r(B)=\begin{cases} m-1, & \text{if } a_2=0 \text{ and } a_1'a_1=1,\\ m, & \text{if } a_2=0 \text{ and } a_1'a_1\neq 1,\\ m+1, & \text{if } a_2\neq 0.\end{cases}$$

B is positive semidefinite iff $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ iff $a_2=0$, $a_1'a_1 \leq 1$, and B is negative semidefinite iff $m=1$, $\lambda_1 \leq 0$, $\lambda_2 \leq 0$ iff $m=1$, $a_2=0$, $a_1'a_1 \geq 1$.

Next let A be a general positive semidefinite matrix of rank m. Then there exists an orthogonal $n\times n$ matrix $S=(S_1:S_2)$ and an $m\times m$ diagonal matrix Λ (with positive diagonal elements) such that $AS_1=S_1\Lambda$, $AS_2=0$. Define the nonsingular $n\times n$ matrix $T=(S_1\Lambda^{-1/2}:S_2)$. Then,

$$T'BT=\begin{pmatrix} I_m & 0\\ 0 & 0\end{pmatrix}-(T'a)(T'a)'$$

and the results follow from the first part of the proof.

REFERENCES

Baksalary, J. K., and R. Kala (1983): "Partial Orderings between Matrices One of which is of Rank One," *Bulletin of the Polish Academy of Sciences*, *Mathematics*, 31, 5–7.

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

——— (1983): "Biased Estimation," in *Handbook of Econometrics*, Vol. I, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland Publishing Company, Chapter 10.

Leamer, E. E. (1978): *Specification Searches*. New York: John Wiley.

Magnus, J. R. (1999a): "The Traditional Pretest Estimator," *Theory of Probability and Its Applications*, forthcoming.

——— (1999b): "Estimation of the Mean of a Univariate Normal Distribution with Known Variance," Manuscript, CentER for Economic Research, Tilburg University.

Magnus, J. R., and J. Durbin (1996): "A Classical Problem in Linear Regression or How to Estimate the Mean of a Univariate Normal Distribution with Known Variance," CentER DP 9660, CentER for Economic Research, Tilburg University.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications*, 2nd Edition. New York: John Wiley.

Toutenberg, H. (1992): *Lineare Modelle*. Heidelberg: Physica-Verlag.

Wold, H. (1953): *Demand Analysis*: *A Study in Econometrics*. New York: John Wiley.