# A NOTE ON APPROXIMATE REGRESSION DISTURBANCES

by

Clifford Hildreth

Discussion Paper No. 2, January, 1971

University of Minnesota

# A NOTE ON APPROXIMATE REGRESSION DISTURBANCES *

Clifford Hildreth

University of Minnesota

## I. Introduction

In a stimulating series of recent articles, Theil [7,8] and others [1, 4, 5, 6] have developed a method of approximating the true disturbances of a linear regression equation and have shown how the approximate disturbances may be used to construct more tractable tests of hypotheses that the true disturbances do conform to our usual assumptions of independence, homoscedasticity, etc.[1]

This work is outlined below with slightly changed emphasis at some points. An alternative approximation is then developed. The alternative is more easily derived and calculated than the approximation proposed in the articles cited. Its accuracy, however, may be either greater or less. Circumstances determining relative accuracy should be explored further along with the properties of statistics based on the alternative approximations.

Let a linear regression model be indicated by

$$y = X\beta + u \tag{1.1}$$

where $y$ is an observed random vector of $T$ components with mean $X\beta$ and variance $\nu I$ with $\nu$ a positive scalar and $I$ an identity matrix

---

* This research was conducted under National Science Foundation grant GS 2286 to the University of Minnesota. Part of the work was performed while the author was a Visiting Professor at Virginia Polytechnic Institute. Useful suggestions were received from a referee and from Marcel K. Richter.

1. Tests of independence against the alternative of a simple autoregressive scheme are the subject of references [1, 4, 5]. Theil [7, p. 1077], [8, p. 247] indicates other possible applications.

of order T. X is a known $T \times K$ matrix of rank K and $\beta$ an unknown vector. u is an unobserved random vector called the disturbance and assumed to have mean 0 and variance $\nu I$.

Theil posed the problem of finding a linear transformation of y that would approximate u in the sense of minimizing the expected value of the sum of squares of differences between corresponding components. He wished the approximate disturbance to be unbiased (have zero mean) and to have a scalar variance matrix (the product of the scalar $\nu$ and an identity matrix).

The problem may thus be initially summarized:

$$\min_{A} E(Ay-u)'(Ay-u) \text{ subject to}$$

$$E(Ay-u) = EAy = AX\beta = 0 \tag{1.2}$$

$$\text{Var } Ay = E\, Ayy'A' = \nu AA' = \nu I \tag{1.3}$$

If A is square and I on the right of (1.3) is of order T, then (1.2) and (1.3) are inconsistent since (1.2) requires $AX = 0$ which means that A and AA' can be of rank at most T-K, while (1.3) requires that A have rank T.

Theil's resolution of the inconsistency is to choose K components of u to be approximated by zero. Let $u_1$ be the subvector of u containing the remaining T-K components to be approximated. The revised problem is

$$\min_{C} E\,(Cy-u_1)'(Cy-u_1) \text{ subject to}$$

$$CX = 0 \tag{1.2*}$$

$$E\, Cyy'C' = \nu CC' = \nu I \tag{1.3*}$$

where I on the right of (1.3*) is of order T-K.

The components of u to be included in $u_1$ are to be chosen with reference to the use to be made of the approximation, see Theil [8, p. 247]. If the components of u are arranged so that the first K are excluded from $u_1$, and y, X, $\tilde{u}$ are correspondingly arranged, then all may be partitioned -

$u = \begin{pmatrix} u_0 \\ u_1 \end{pmatrix}$ , $y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$ , $X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}$ , $\tilde{u} = \begin{pmatrix} \tilde{u}_0 \\ \tilde{u}_1 \end{pmatrix}$ where $\tilde{u}$ is the vector of least-squares residuals of the regression of y on X.

Alternative expressions for the least-squares residuals are -

$$\tilde{u} = y - X\tilde{\beta} = (I - X(X'X)^{-1}X')\,y = My = Mu \tag{1.4}$$

where $\tilde{\beta} = (X'X)^{-1}X'y$ are the least-squares estimates of $\beta$ and M is seen to be a symmetric, idempotent matrix of rank T-K. Theil shows that the minimizing approximations, say $\hat{u}_1$, may be written

$$\hat{u}_1 = \tilde{u}_1 - X_1 X_0^{-1} \left[ \sum_{k=1}^{K} \frac{d_k}{1+d_k} q_k q_k' \right] \tilde{u}_0 \tag{1.5}$$

where the $d_k$ are positive square roots of the eigenvalues of $X_0(X'X)^{-1}X_0'$ and the $q_k$ are the corresponding latent vectors.

Before turning to an alternative approximation let me note that I do not think the criterion of unbiasedness is very important in this context. If the expectation of the sum of squared differences between components of u and the approximating vector (approximator) could be materially reduced by waiving this requirement it would seem to me to be desirable to accept some bias.

However, in the present context this possibility does not exist. If the approximator is to be linear in y, then it is seen from equation (1.2) above that the condition necessary for the approximator to have mean zero is the same as the condition that it not depend on the unknown $\beta$-namely, the transformation applied to $y$ must annihilate $X$. Such a requirement will be maintained in what follows, but I am motivated by the necessity for avoiding an approximator that depends on $\beta$ (except through $y$) rather than wanting to impose unbiasedness.

## II. An Alternative Approximator

Following Theil, we want to find a random vector, call it $w$, that will approximate the unknown disturbance $u$ in the sense that $E(w-u)'(w-u)$ is small. $w$ cannot depend on $\beta$ (we wish to be able to calculate $w$ in particular applications) and it is desired that $\text{Var } w = \nu I$.

As Theil observed, it seems natural to start by considering linear transformations of the observed vector $y$, say $Ay$. As noted in the first section, if $Ay$ is not to depend on $\beta$, it is necessary that

$$AX = 0 \tag{2.1}$$

which means that the rank of $A$ (denoted $\rho A$) can be at most $T-K$ (since $\rho X = K$).

For $w$ to have a nonsingular variance, the loss of rank imposed by (2.1) must be restored. One possibility is to add $K$ additional observable random variables to $y$ as possible ingredients of $w$.

To explore such an approach, let $\epsilon$ be a standard normal vector of order $K$. Let $\bar{\nu} = \frac{1}{T-K} \tilde{u}'\tilde{u}$ be the usual unbiased estimator of $\nu$. Define

$$\bar{\nu}^{\frac{1}{2}} \epsilon = r \tag{2.2}$$

Then $E\, r = 0$, $E\, rr' = \nu I$. In applying this approach, an observed value of $\epsilon$ would be drawn from an appropriate table of random numbers.

Let $B$ be a matrix of order $T \times K$. Consider an approximator of the form

$$w = Br + Ay \tag{2.3}$$

where $A$ and $B$ are chosen to minimize $E(w-u)'(w-u)$ subject to (2.1) above and

$$\text{Var } w = Eww' = \nu(BB' + AA') = \nu I \tag{2.4}$$

or

$$BB' + AA' = I \tag{2.5}$$

In the remainder of this section it is shown that the restricted minimum is achieved for

$$A = M = (I-X(X'X)^{-1}X') \quad \text{and} \tag{2.6}$$

$$B = X(X'X)^{-\frac{1}{2}} \quad \text{and} \tag{2.7}$$

these values are unique except for alternative square roots of $(X'X)^{-1}$. Consequences of the results are briefly discussed in a concluding section. To establish the results, two lemmas are useful.

<u>Lemma 1</u>. If $B$ is a matrix of order $T \times K$ with $K < T$ and if $A$ is a $T \times T$ matrix of rank $\leq (T-K)$ such that $BB' + AA' = I$, then

(1) $BB'$ is idempotent and of rank $K$

(2) $AA'$ is idempotent and of rank $T-K$.

<u>Proof</u>: Let $G$ be an orthogonal matrix that diagonalizes $BB'$, say $GBB'G' = D_1$. Then

$$GBB'G' + GAA'G' = GG' \quad \text{or}$$

$$D_1 + GAA'G' = I$$

so $G$ also diagonalizes $AA'$. Let $GAA'G' = D_2$,

$$D_1 + D_2 = I$$

$D_1$ has at most $K$ nonzero elements because $B$ has $K$ columns. $D_2$ has at most T-K nonzeros since $\rho A \leq (T-K)$. But $D_1 + D_2$ has T nonzeros so

$$\rho D_1 = \rho BB' = \rho B = K \quad \text{and} \quad \rho D_2 = \rho AA' = \rho A = T-K$$

Since each nonzero of $D_1$ and $D_2$ must be unity, $BB'$ and $AA'$ are idempotent.

Lemma 2. Let X be a given matrix of order $T \times K$ and rank K. Let $\mathfrak{X}$ be the subspace of $R^T$ spanned by the columns of X. Let $\mathcal{Z}$ be the subspace of $R^T$ complementary to $\mathfrak{X}$ ($\mathfrak{X} \oplus \mathcal{Z} = R^T$, $\mathfrak{X} \cap \mathcal{Z} = \{0\}$). Let Z be a given $(T-K) \times T$ matrix whose rows form an orthonormal basis for $\mathcal{Z}$($ZZ' = I$, $ZX = 0$). If H is any $T \times (T-K)$ matrix satisfying $\mathrm{tr}\, HH' = T-K$, then

(3) $\mathrm{tr}\, HZ \leq T-K$

(4) $\mathrm{tr}\, HZ = T-K \Leftrightarrow H = Z'$

Proof: Consider the problem of maximizing $\mathrm{tr}\, HZ$ subject to $\mathrm{tr}\, HH' = T-K$. This is a simple problem of maximizing a linear function of $T(T-K)$ variables (the elements of H) subject to the restriction that the values lie on a sphere of radius $\sqrt{T-K}$ in $R^{T(T-K)}$. The Lagrangian is

$$L(H,\lambda) = \sum_{t=1}^{T} \sum_{s=1}^{T-K} h_{ts} z_{st} + \lambda \sum_{t=1}^{T} \sum_{s=1}^{T-K} h_{ts}^2$$

$$\frac{\partial L}{\partial h_{ts}} = z_{st} + 2\lambda\, h_{ts} \qquad t = 1,2,\ldots,T; \quad s = 1,2,\ldots T-K$$

so the maximizing values $\bar{h}_{ts}$ must satisfy.

$$\bar{h}_{ts} = -\frac{1}{2\lambda} z_{st} \quad \text{or} \quad \bar{H} = -\frac{1}{2\lambda} Z' .$$

Therefore

$$\operatorname{tr} \ \overline{H}\overline{H}' = \frac{1}{4\lambda^2} \operatorname{tr} ZZ' \quad \text{or} \quad T\text{-}K = \frac{1}{4\lambda^2} (T\text{-}K) \ .$$

Hence

$$\lambda = \pm \tfrac{1}{2} \ .$$

Noting that the constraint set is the boundary of a strictly convex, bounded set we conclude that $\overline{H} = Z'$ corresponds to a unique restricted maximum and $\overline{H} = -Z'$ to a unique restricted minimum.

The desired results readily follow in

<u>Theorem 1</u>. Let $w$ be a vector of the form given in (2.3) with $r,y$ as defined in (2.2), (1.1). Then $E(w-u)'(w-u)$ is minimized over $B,A$ subject to

(a) $AX = 0$

(b) $BB' + AA' = I$

by

$$A = M = (I-X(X'X)^{-1}X') \qquad \text{and} \tag{2.6}$$

$$B = X(X'X)^{-\frac{1}{2}} \tag{2.7}$$

and the minimizing values are unique except for alternative values of $(X'X)^{-\frac{1}{2}}$ .

<u>Proof</u>: From (1.1), (2.3) and Condition (a), $w = Br + Au$. Since $u$ and $r$ are uncorrelated, $E(w-u)'(w-u) = Ew'w - 2Ew'u + Eu'u = \nu \operatorname{tr}(B'B + A'A - 2A + I)$. Since $\operatorname{tr}(B'B + A'A) = \operatorname{tr}(BB' + AA') = T$ by Condition (b), this reduces to $2\nu(T - \operatorname{tr} A)$. Thus, the problem is equivalent to minimizing $\operatorname{tr} A$ subject to (a) and (b).

Condition (a) requires that the rows of $A$ be orthogonal to the columns of $X$, so $A$ is of the form

$$A = HZ$$

where the rows of $Z$ form an orthonormal basis for the subspace $\mathcal{Z}$ of $R^T$ complementary to the subspace spanned by the columns of $X$. By Lemma 1, Condition (b) implies $\operatorname{tr} AA' = T-K$. Therefore the conditions of Lemma 2 hold and $H = Z'$ is the unique maximizing value (for $\operatorname{tr} A$, minimizing for $E(w-u)'(w-u)$ for a given $Z$). However, if $W$ is an alternative matrix whose rows are an orthonormal basis for $\mathcal{Z}$, then $W = PZ$ with $P$ orthogonal[2] and $W'W = Z'P'PZ = Z'Z$ so $A$ is uniquely determined and $\operatorname{tr} A = \operatorname{tr} Z'Z = T-K$.

Now $M = I-X(X'X)^{-1}X'$ clearly satisfies the conditions of the theorem and $M$ is known to be idempotent and rank $T-K$. Therefore $\operatorname{tr} M = T-K$ so $A = M$ is the unique restricted minimizer of $E(w-u)'(w-u)$. From (b), $BB' = I-AA' = I-M = X(X'X)^{-1}X'$.

## III. Some Observations

From Theorem 1, the approach of Section II leads to an approximator of the form

$$\begin{aligned} w &= X(X'X)^{-\frac{1}{2}}r + My \\ &= \sqrt{\nu}\, X(X'X)^{-\frac{1}{2}}\varepsilon + \tilde{u} \end{aligned} \tag{3.1}$$

where $\varepsilon$ is a drawing from a standard multivariate normal population of order $K$ and $\tilde{u}$ is the vector of least-squares residuals from the regression of $y$ on $X$. Using $E(w-u)'(w-u)$ as an indicator of the

2. Clearly $W=PZ$ for some $P$ since $Z$ is a basis. Since $I = WW' = PZZ'P' = PP'$, $P$ must be orthogonal.

inaccuracy of the approximation and using (1.4) to write
$w-u = \sqrt{\bar{\nu}}\, X(X'X)^{-\frac{1}{2}}\epsilon + (\tilde{u} - u) = \sqrt{\bar{\nu}}\, X(X'X)^{-\frac{1}{2}}\epsilon + X(X'X)^{-1}X'u$, we have

$$E(w-u)'(w-u) \doteq \nu \operatorname{tr} X(X'X)^{-1}X' + \nu \operatorname{tr} X(X'X)^{-1}X' = 2\nu K \qquad (3.2)$$

Inaccuracy of $\nu K$ is incurred by having to observe $\tilde{u}$ instead of $u$, and another $\nu K$ is incurred to rectify the second moment of $\tilde{u}$.

Calculations using Theil's approach, see [4, pp. 179-80] and [7, pp. 1075-6], show that it is sometimes more and sometimes less accurate than the approach developed here. Relative accuracy should be investigated further and, more important, the behavior of the statistics based on the alternative approximations should be studied. Comparisons of the power of tests of the hypotheses of serial independence and homoscedasticity of disturbances will be of particular interest.[3]

Tests based on $w$ will be approximate since $w$ is not exactly normally distributed. It seems reasonable to conjecture that when the sample is sufficiently large that the distribution of $\bar{\nu}$ is pretty concentrated, the distribution of $w$ should be approximately normal. $w$ is a probability mixture of two multivariate normal vectors. In a different but analogous context,[4] the conjecture of approximate normality was later supported by results of a Monte Carlo study.

It appears then that the approach suggested here has the advantage of simplicity as compared with Theil's approach, but simplicity is purchased at the sacrifice of exact normality (in those cases in which $u$ is exactly normal). The more important issues of accuracy of

---

3. For certain special cases comparisons of the relative power of Theil's test and the Durbin-Watson test are included in [1] and [5].
4. See [2], section 5, and [3], pp. 25-6.

approximation and power of tests need to be explored in more detail since relative accuracy, and presumably power, depends on the nature of the matrix $X$ of observed values of independent variables in particular applications.

## References

[1] Abrahamse, A. P. J., and Koerts, J., "A comparison between the power of the Durbin-Watson test and the power of the BLUS test." Journal of the American Statistical Association, 64 (1969), 938-948.

[2] Hildreth, Clifford, "Asymptotic distribution of maximum likelihood estimators in a linear model with autoregressive disturbances," Annals of Mathematical Statistics, 40, (1969), 583-94.

[3] Hildreth, Clifford, and Lu, John Y., "A Monte-Carlo study of the regression model with autocorrelated disturbances," Rand Memorandum RM5728PR, Santa Monica, 1969.

[4] Koerts, J., "Some further notes on disturbance estimates in regression analysis." Journal of the American Statistical Association, 62 (1967) 169-183.

[5] Koerts, J. and Abrahamse, A. P. J., "On the power of the BLUS procedure," Journal of the American Statistical Association, 63 (1968), 1227-1236.

[6] Neudecker, "A Note on BLUS Estimation," Journal of the American Statistical Association, 64 (1969), 949-952.

[7] Theil, H., "The analysis of disturbances in regression analysis," Journal of the American Statistical Association, 60 (1965), 1067-1079.

[8] Theil, H., "A simplification of the BLUS procedure for analyzing regression disturbances," Journal of the American Statistical Association, 63 (1968), pp. 242-251.