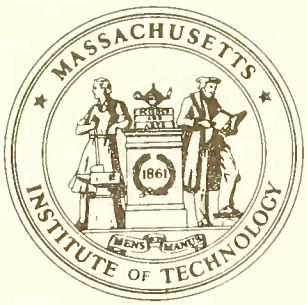


working paper
department
of economics

SOME FINITE SAMPLE PROPERTIES OF SPECTRAL ESTIMATORS OF A LINEAR REGRESSION

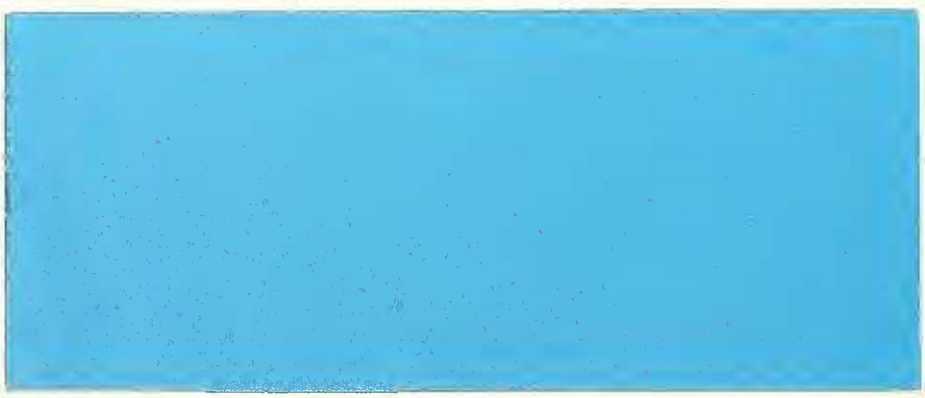
by

Robert F. Engle and Roy Gardner[1]

Number 122 December 1973

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139



SOME FINITE SAMPLE PROPERTIES OF SPECTRAL ESTIMATORS OF A LINEAR REGRESSION

by

Robert F. Engle and Roy Gardner[1]

Number 122 December 1973

The authors are indebted to the National Bureau of Economic Research Computer Research Center for provision of computer time and use of the TROLL system and to the Cambridge Project for additional support. The work was undertaken while Engle was on leave at Cornell University.

1 The authors are from Massachusetts Institute of Technology and Cornell University respectively.

SOME FINITE SAMPLE PROPERTIES OF SPECTRAL ESTIMATORS OF A LINEAR REGRESSION

by

Robert F. Engle and Roy Gardner

In this paper we consider estimation of a linear regression with covariance stationary disturbances:

$$(1) \qquad y = x\beta + \varepsilon \quad E\,\varepsilon\varepsilon' = \sigma^2\Omega \quad E(\varepsilon \mid x) = 0.$$

If consistent estimates of the elements of Ω are available, then subject to some regularity conditions on x, generalized least squares based on the estimated covariance matrix will be asymptotically efficient. If an inconsistent estimate of this matrix is used, then the estimator may be asymptotically inefficient depending upon the process of the exogenous variables.

Although Ω has T unknown parameters, where T is the length of the time series, these can be estimated consistently in either the time domain or the frequency domain by parameterizing the matrix by a number of parameters which increases with T. The classic estimator is Hannan's (9), (10), spectral estimator which approximates Ω by a transformation of the estimated spectrum of ε. Amemiya (1) approximates Ω by an estimated nth order autoregression on ε where n grows with the sample although he does not estiblish the rate. Both of these estimators are asymptotically equivalent to the best linear unbiased estimator, and are asymptotically efficient. In this paper we focus on Hannan's procedure although one version is quite similar to Amemiya's estimator.

These asymptotically efficient estimators are not often used; instead, it is common to approximate the process of the disturbance by a low order autoregressive process and then estimate as if the approximation were exact. If it is not, and this inconsistent estimate of Ω is used in the GLS procedure,

then the estimate will generally be asymptotically inefficient. Watson and Hannan (14) early evaluated the loss of efficiency and showed that it can be very substantial. Engle (6) has extended this discussion to show that often, even OLS is asymptotically superior to a low order autoregressive approximation. That is, a feeble attempt to correct for serial correlation is not necessarily better than none at all.

There are two general explanations for the neglect of Hannan's estimator in empirical studies in the face of its theoretical attractiveness. First, it is widely believed, although not proven, that the small sample properties of the estimator are far inferior to its asymptotic behavior, and that for relevant sample sizes, a low order approximation to the error process is good enough. Second, the computational burden of even Hannan's version is considered sufficient to outweigh the gains in efficiency for the casual investigator. In this paper we investigate the validity of these beliefs.

We can view Hannan's estimator as a non-parametric approach to dealing with dependence among the errors of a linear regression. As such, it is robust against a variety of misspecifications of the error process without costing anything asymptotically. The important cost must be for finite samples and a better understanding of the trade-off between robustness and finite sample properties at economically relevant sample sizes would aid investigators in making this choice. To describe this trade-off, we estimate the second order term in the asymptotic expansion for the variance of each of the estimators.

In section II we derive our version of Hannan's estimator very simply; in sections III and IV the estimators and stochastic environments are described. Section V presents the detailed results while section VI gives the pooled results and, in particular, our estimate of the second order term in the expansion for the asymptotic variance of all the estimators. Section VII summarizes the implications of the study.

II. The Hannan Estimator

In the context of (1) the GLS estimator is

$$\hat{\beta} = (x'\Omega^{-1}x)^{-1}x'\Omega^{-1}y \quad (2)$$

We define the matrix W by its typical element

$$W_{st} = \exp(2\pi i/T)st \qquad s, t = 0, \ldots, T-1 \quad (3)$$

which can be shown to be unitary[1] and which produces a finite fourier transform of a time domain vector. Rewriting (2) we obtain

$$\hat{\beta} = (\tilde{x}' A \tilde{x})^{-1}\tilde{x}' A \tilde{y} \quad (4)$$

where $\tilde{x} = Wx$ is the finite fourier transform of x and

$$A = W'\Omega^{-1}W. \quad (5)$$

If either the error process is a circulant, or the number of observations is large, the matrix A will be diagonal with elements which are the inverse of the spectrum of the disturbances at the harmonic frequencies. Hannan's estimator approximates A by a diagonal matrix $\hat{A}$ which has as elements, the inverse of a consistent estimate of the spectrum of the disturbances. This is therefore asymptotically equivalent to generalized least squares and is well known to be consistent and asymptotically efficient, under rather general conditions. Just as in the time domain versions (12), (3), this proof requires that there be no lagged dependent variables among the x's.

---

[1] We will use $i = \sqrt{-1}$. Because W is a complex matrix we shall throughout interpret a prime as the complex conjugate of the transpose. For real variables this of course has no effect.

Several variations of this estimator are possible depending upon which estimate of the disturbance spectrum is employed. Asymptotically this will make no difference while for finite samples it may.[2] Once the choice of $\hat{A}$ has been made a simple computational procedure is to define $x^* = W'\hat{A}^{-1/2}Wx$ and $y^* = W'\hat{A}^{-1/2}Wy$ and then use ordinary least squares on these transformed variables. For more discussion of these procedures and related applications, see Engle (5).

Rewriting (4) in a form more familiar to frequency domain analysts with $I_{xx}(s)$ as the matrix of cross-periodograms of all the x's at frequency s, we obtain:

$$\hat{\beta} = \left[\sum_s I_{xx}(s)\,\hat{A}(s)\right]^{-1} \sum_s I_{xy}(s)\hat{A}(s) \quad . \tag{6}$$

The most common types of spectral estimates can all be viewed as a weighted average of adjacent elements of the periodogram. In this case letting u be the residuals from a first stage consistent, but not efficient, estimate of (1), we can write our estimator as

$$\hat{\beta} = \left[\sum_s I_{xx}(s)/\left(\sum_r I_u(s-r)\xi(r)\right)\right]^{-1} \sum_s I_{xy}(s)/\left(\sum_r I_u(s-r)\xi(r)\right) \tag{7}$$

where the choice of the estimator is merely the choice of a set of weights ξ.

This estimator is not identical with that proposed by Hannan (10) although we will show that the differences are unimportant. Hannan's formulation replaces all the periodograms in (7) by consistent estimates of the spectra of the x's and y. That is, in place of the periodogram of x is a weighted average of adjacent periodograms of x. Letting these weights be given

---

2 This is parallel to Griliches and Rao (8) who investigated several approaches to estimation of a first order Markov model which differed only for finite samples.

by ϕ and ψ, we can write Hannan's estimator as:

$$(8)\quad \hat{\beta} = \sum_s \sum_q I_{xx}(s-q)\phi(q)) / \left(\sum_r I_u (s-r)\psi(r)\right)\Big]\Big\}^{-1} \sum_s \Big[\Big(\sum_q I_{xy}(s-q)\phi(q)\Big) \Big/ \Big(\sum_r I_u (s-r)\psi(r)\Big)$$

Grouping terms however we obtain

$$(9)\quad \hat{\beta} = \Big\{\sum_s I_{xx}(s) \Big[\sum_q \phi(q) / \sum_r I_u (s-q-r)\psi(r)\Big]\Big\}^{-1} \sum_s I_{xy}(s) \Big[\sum_q \phi(q) / \Sigma I_n (s-q-r)\psi(r)\Big]$$

Comparing (7) and (9) we see that one is a weighted average of the inverse of a weighted average, while the other is merely the inverse of a weighted average. There is no non-trivial way that the weights can be chosen to make these identical; however, when the denominator changes little, as would be the case when the sample is large so that very close frequencies are being averaged, the difference between the two procedures could always be made quite small through choice of ϕ and ψ. Since we know so little about optimal weights for finite samples in the regression context, it seems unlikely that one formulation can be shown to be statistically superior in any way to the other. The advantage of (7) is that only one set of weights need be chosen and it is computationally easier not to smooth the other periodograms.

III. Choice of Estimators

For the tests of Hannan's estimator, three versions were implemented. Each takes the residuals from an OLS regression of the model and estimates the spectrum of the disturbances.

HAN 1. We take a symmetrical rectangular moving average of the periodogram of the residuals as the estimate of the disturbance spectrum. This has the advantage of having the minimum mean squared error if the true process is white noise. It has the disadvantage that near the endpoints of the

spectrum the variance becomes much larger because the moving average covers fewer elements. Because economic exogenous variables have so much spectral weight at low frequencies (7), a large variance in the residual spectrum estimate at this point may mean a very noisy estimator. For sample sizes 50, 100 and 200, bandwidths of .07, .05 and .035 were used.

HAN 2. We assume that the disturbance spectrum is piecewise constant. Thus the average over a series of frequencies is attributed to all these frequencies and the spectrum looks like a series of steps. This has been used before by Duncan and Jones (4), and aside from the obvious computational advantages, does give a smaller variance estimate of the low frequency spectrum than Type 1. It of course does not capture the fine shape of the spectrum which again may be important at low frequencies. For samples of size 50, 100, and 200 we used 7, 10 and 14 intervals respectively.

HAN 3. We assume that that the disturbance process is an n^{th} order auto-regression. Therefore, by regressing the residual vector on n lagged values of itself, the parameters of this process are consistently estimated. These parameters imply the following spectrum:

$$(10) \qquad f_u(\Theta) = 1/\left| 1 - \gamma_1 e^{i\Theta} - \gamma_2 e^{2i\Theta} - \gamma_3 e^{3i\Theta} - \ldots - \gamma_n e^{ni\Theta} \right|^2$$

where the γ's are the estimated autoregressive coefficients. This procedure is somewhat more parametric since the number of unknown parameters in the spectrum can be easily restricted. As the number of observations becomes very large, the choice of n should similarly be increased. This method has recently been recommended by Parzen (14) for spectrum estimation. In our case it has the further advantage that as long as n is large enough, the structure of the spectrum at low frequences is no more difficult to

discern than at any other frequency. For all runs we chose n = 2.

ALS. We assume that the error process is generated by a first order Markov process and estimate the model using maximum likelihood procedures which we name autoregressive least squares. This well known and widely used estimator is included to provide a comparison with the more unusual spectral estimates.

OLS. Ordinary least squares is always a standard of comparison and its robustness in new situations is generally impressive.

IV. Stochastic Environments

The equation which was simulated and estimated repeatedly was

$$y = \alpha + x\beta + \varepsilon \tag{11}$$

$$\varepsilon = \gamma_1 \varepsilon_{-1} + \gamma_2 \varepsilon_{-2} + \eta$$

$$\sigma^2_\varepsilon = 1/4, \Sigma(x_t - \bar{x})^2 / (T - 1) = 1$$

$$\alpha = 0, \quad \beta = 1$$

For each of five assumptions about the error process, five different x processes, and three sample sizes, a set of ten realizations were calculated, for a total of 750 independent data sets. On each of these data, the five estimators described above were evaluated. Relatively few realizations were calculated for each environment on the grounds that more information would be obtained by pooling over a variety of cases.

The x processes were chosen to represent typical situtations for eocnomic time series analysis. Five economic time series were used to identify typical processes. These series were quarterly constant dollar gnp, quarterly current dollar corporate profits, quarterly seasonally unadjusted expenditures on plant and equipment, monthly Standard and Poors

stock price index, and monthly wholesale price index, seasonally unadjusted. Each was logged[1], detrended with a linear function of time, and fitted by a fourth order autoregression. The fits were all quite good, and we can think of these as Box-Jenkins models (2). When we solved for the roots of these fourth order processes, all were less than one in absolute value[2] and many were complex, indicating oscillating components. Table 1 gives the parameters of these processes.

The error processes chosen for the test were picked to give a wide range of experience for the estimators. They range from white noise and first order markov processes to highly dependent oscillating and damped second order processes. Their specific defining characters are given in Table II, as well as their spectral shapes. From Engle (6) we know that it is likely that OLS will dominate ALS for error 5 because of the change of slope at low frequency, and that it is possible that the same will be true for error 3 where the very steep slope cannot be approximated by a first order process. Furthermore, for error 4, OLS and ALS should be equal and inferior to GLS since the first order serial correlation coefficient is so small.

Three independent samples of sizes 50, 100 and 200 were drawn for each environment. These were picked to represent the common range of post-war economic quarterly and monthly data. To minimize initial value problems, the first 200 observations drawn[3] were discarded in each case. The remainder

---

1 The stock price index was left in levels.

2 However, when these were estimated in levels, rather than logs, all except stock prices were unstable and therefore appeared non-stationary.

3 The random numbers were computed by the SNORM subroutine of the TROLL system, which essentially uses a polar transformation of uniform variates to obtain normal random variables. The uniform numbers are obtained from a linear bicongruential generator which truncates a set of large numbers and then uses a second set to randomly shuffle the first, thereby eliminating all possibilities of serial dependence. These procedures are described in Knuth (11, Chapter 3).

TABLE 1

X PROCESSES

$$(1 - \gamma_1 L - \gamma_2 L^2 - \gamma_3 L^3 - \gamma_4 L^4)X_t = \varepsilon_t$$

$$(1 - \alpha_1 L)(1 - \alpha_2 L)(1 - \alpha_3 L)(1 - \alpha_4 L)X_t = \varepsilon_t$$

VARIABLE	γ_1	γ_2	γ_3	γ_4	$\lvert\alpha_1\rvert$	$\lvert\alpha_2\rvert$	$\lvert\alpha_3\rvert$	$\lvert\alpha$
GNP, Quarterly	1.343	-.182	-.303	.117	.829	.592*	.551	
PR, Quarterly	1.144	-.186	-.005	-.093	.766*	.300*		
I, Quarterly, Not Seasonally Adjusted	.081	.026	-.207	.720	.966	.918*	.885	
SP, Monthly	1.200	-.300	.150	-.100	.924	.550	.443*	
WPI, Monthly, Not Seasonally Adjusted	1.355	-.218	-.142	.004	.975	.667	.214*	

* A pair of complex roots

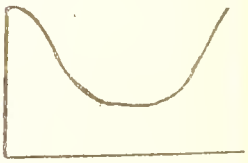
TABLE II

ERROR PROCESSES: U

$$(1 - \gamma_1 L - \gamma_2 L^2)U_t = (1 - \delta_1 L)(1 - \delta_2 L)U_t = \varepsilon_t$$

$$\rho = \frac{\delta_1 + \delta_2}{1 + \delta_1\delta_2} = \frac{\gamma_1}{1 - \gamma_2}$$

$$\text{SLOPE} = (1 - \delta_1)^{-2}(1 - \delta_2)^{-2}\left(\frac{\delta_1}{(1 - \delta_1)^2} + \frac{\delta_2}{(1 - \delta_2)^2}\right)$$

ERROR	γ_1	γ_2	δ_1	δ_2	ρ	SLOPE	SPECTRUM
1.	0.	0.	0.	0.	0.	0.	
2.	.6	0.	.6	0.	.6	23.4	
3.	1.7	-.72	.9	.8	.988	275,000.	
4.	.1	.72	.9	-.8	.357	2777.0	
5.	1.4	-.98	$(.7 \pm .7i)$		.707	10.14	

were tested for normality and time dependence.

All computation was done on the National Bureau of Economic Research TROLL time sharing computer system which uses an IBM 360-67 computer. Although most of the programming is in single precision, the important fast fourier transform subroutine is calculated in double precision. The estimators are available for general use through this system.

V. Detailed Results

The choice of an estimator should be made on the grounds of mean squared error if the investigator has a quadratic loss function. That is, we should compare the sum of squared bias and variance across estimators. We found that the bias term was very small and conclude that all the estimators are essentially unbiased at these sample sizes.

To obtain this result, we computed the ratio of the bias to the standard deviation of the parameter estimates for each of the 375 environments. Since each case is 10 independent observations on a normal (by construction) parameter estimate, this ratio should be distributed as t with 9 degrees of freedom times root 10 for an unbiased estimator. Only 4 ratios exceeded the 95% critical value of .715 which is well below the expected 5%. These all occurred for the estimator HAN 3 and all are negative, however, all occurred for a sample of 200 observations while for the same environment, the ratios for smaller samples were quite small and frequently positive. Furthermore, these were observations with especially small standard deviations, not large biases. Thus, while it is conceivable that we observe a bias in HAN 3, it does not behave like a small sample bias and we shall ignore it as random fluctuation.

The error we commit by examining variance throughout the paper rather than mean squared error is to underestimate the error by one plus the square of this ratio. Since this is rarely as high as one half, we are generally making much less than a 25% error by ignoring the bias.

The variances of all the estimators in each environment are given in Tables III-VII. Each entry is a sample variance of 10 observations and therefore the confidence interval associated with any particular number is rather large. Specifically, the 90% confidence interval covers all values between one half and two and a half times the estimated value. Most of the interesting differences between estimators are large and by pooling over environments we substantially shrink the confidence intervals.

In each case the asymptotic variance for a sample of 200 observations is presented for comparison. Since the stochastic processes are known for each situation, the asymptotic variance can be calculated by numerical integration of the relevant spectral density functions.

$$(12) \qquad T\,AVAR\,(\hat{\beta}_{OLS}) = \sigma^2 \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta)d\theta\right]^{-2} \frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta)\, f_u(\theta)\, d\theta$$

$$(13) \qquad T\,AVAR\,(\hat{\beta}_{ALS}) = \sigma^2 \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta)\, f_{\underline{u}}^{-1}(\theta)\, d\theta\right]^{-2} \frac{1}{2\pi}\int_{\pi}^{\pi} f_x(\theta) f_u(\theta) f_{\underline{u}}^{-2}(\theta)d\theta$$

$$(14) \qquad T\,AVAR\,(\hat{\beta}_{HAN}) = \sigma^2 \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta)\, f_u^{-1}(\theta)d\theta\right]^{-1}$$

As a convention we normalize x and u to obtain

$$(15) \qquad \frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(\theta)d\theta = 1 = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi} f_u(\theta)d\theta\,, \quad \sigma^2 = 1/4$$

The spectra of x and u are given by (10) with the appropriate set of γ's, and the spectrum of $\underline{u}$, which is the truncated first order approximation. is given by (10) where n = 1 and γ_1 is the true first order serial correlation coefficient.

Error process 1 is just white noise and therefore all estimators are asymptotically equivalent. Notice that for samples of size 50, OLS is minimum variance (indicated by a star) for four out of five x processes.

TABLE III

TABLE OF VARIANCES ($\times 10^5$)

ERROR = 1

X PROCESS	T	OLS	ALS	HAN 1	HAN 2	HAN 3
GNP, Q	50	620.	630.	500.*	940.	640.
	100	240.	240.	320.	270.	230*
	200	160.	160.	120.*	140.	160.
	AVAR†	130.	130.	130.	130.	130
PR, Q	50	1000.*	1100.	1100.	1300.	1100.
	100	250.	250.*	310.	320.	260.
	200	160.	160.	120.*	130.	160.
	AVAR†	130.	130.	130.	130.	130.
I, Q, NSA	50	730.*	790.	1100.	920.	760.
	100	230.	220.*	250.	310.	240.
	200	170.	160.	120.	120.*	160.
	AVAR†	130.	130.	130.	130.	130.
SP, M	50	550.*	560.	560.	650.	570.
	100	170.	160.	210.	170.	160.*
	200	160.	160.	140*	140.	160.
	AVAR†	130.	130.	130.	130.	130.
WPI, M, NSA	50	580.*	600.	670.	600.	590.
	100	70.	80.	110.	60.*	80.
	200	90.	90.	90.	90.*	90.
	AVAR†	130.	130.	130.	130.	130

* Best estimator for this environment and sample size

† Asymptotic variance calculated for T = 200.

TABLE IV

TABLE OF VARIANCES ($\times 10^{5}$)

ERROR = 2

X PROCESS	T	OLS	ALS	HAN 1	HAN 2	HAN 3
GNP, Q	50	2300.	2000.	1800.*	2900.	2000.
	100	630.	590.	750.	570.*	600.
	200	610	550.	410.*	440.	560.
	AVAR†	410.	340.	340.	340.	340.
PR, Q	50	2700.	1800.*	2100.	2400.	1900.
	100	430.	390.*	440.	410.	410.
	200	490.	320.	240.	230.*	340.
	AVAR†	340.	230.	230.	230.	230.
I, Q, NSA	50	300.	310.	290.*	290.	290.
	100	200.	150.	200.	160.	140.*
	200	170.	50.	60.	50.*	50.
	AVAR†	100.	50.	50.	50.	50.
SP, M	50	2100.	1400.*	1500.	1700.	1600.
	100	590.	480.	620.	390.*	430.
	200	660.	580.	510.	500.*	580.
	AVAR†	450.	380.	380.	380.	380.
WPI, M, NSA	50	2100.	1500.	1900.	1520.*	1700.
	100	420.	310.	420.	180.	210.
	200	400.	430.	420.	450.	440.
	AVAR†	500.	490.	490.	490.	490.

* Best estimator for this environment and sample size

† Asymptotic variance calculated for T = 200.

TABLE V

TABLE OF VARIANCES ($\times 10^5$)

ERROR = 3

X PROCESS	T	OLS	ALS	HAN 1	HAN 2	HAN 3
GNP, Q	50	4400.	40.*	440.	330.	830.
	100	1600.	40.*	100.	70.	110.
	200	1100.	20.*	190.	40.	60.
	AVAR †	1100.	40.	10.	10.	10.
PR, Q	50	2100.	20.*	220.	110.	110.
	100	1000.	30.	60.	40.	10.*
	200	500.	10.	50.	20.	8.5*
	AVAR †	500.	10.	3.9	3.9	3.9
I, Q, NSA	50	270.	1.6*	30.	10.	2.1
	100	540.	1.2*	7.9	4.7	2.4
	200	200.	1.7	1.4	0.6	2.5
	AVAR †	200.	2.3	2.0	2.0	2.0
SP, M	50	6200.	180.*	1900.	2000.	1700.
	100	2300.	50.*	250.	220.	290.
	200	2200.	50.*	260.	220.	130.
	AVAR†	2100.	60.	20.	20.	20.
WPI, M, NSA	50	6100.	550.*	4200.	3500.	3300.
	100	2700.	250.*	750.	690.	740.
	200	2200.	430.*	960.	1200.	1100.
	AVAR†	3300.	900.	190.	190.	190.

* Best estimator for this environment and sample size

† Asymptotic variance calculated for T = 200.

TABLE VI

TABLE OF VARIANCES ($\times 10^{5}$)

ERROR = 4

X PROCESS	T	OLS	ALS	HAN 1	HAN 2	HAN 3
GNP, Q	50	5400.	4300.	2500.	2700.	2200.*
	100	860.	710	320.	360.	260.*
	200	900.	430.	310.	270.	240.*
	AVAR †	610.	520.	180.	180.	180.
PR, Q	50	2300.	2000.	1100.	940.*	1400.
	100	480.	410.	170.	150.*	190.
	200	390.	180.	120.	100.	100.*
	AVAR †	330.	250.	90.	90.	90.
I, Q, NSA	50	2000.	2300.	1900.	1300.	740.*
	100	590.	620.	220.	130.	50.*
	200	290.	290.	80.	70.	70.*
	AVAR †	300.	330.	50.	50.	50.
SP, M	50	3800.	3600.	2700.	3000.	2600.*
	100	1000.	1100.	500.	380*	380.
	200	1300.	1100.	500.*	490.*	460.*
	AVAR †	1000.	930.	300.	300.	300.
WPI, M, NSA	50	3800.	3700.	3300.	3000.	3000.*
	100	1100.	1100.	360	410.	350.*
	200	1100.	1000.	840.	730.*	770.
	AVAR †	1500.	1500.	1100.	1100.	1100.

* Best estimator for this environment and sample size

† Asymptotic variance calculated for T = 200.

TABLE VII

TABLE OF VARIANCES ($\times 10^5$)

ERROR = 5

X PROCESS	T	OLS	ALS	HAN 1	HAN 2	HAN 3
GNP, Q	50	2100.	2600.	730.	200.	130.*
	100	2300.	4600.	250.	90.	60.*
	200	670.	1400.	30.	10.	9.4*
	AVAR†	200.	1600.	7.3	7.3	7.3
PR, Q	50	430.	650.	40.*	50.	110.
	100	600.	1400.	40.*	30.	10.*
	200	130.	340.	6.8	3.4*	8.4
	AVAR†	420.	1100.	5.4	5.4*	5.4
I, Q, NSA	50	40.	6.6	8.7	2.0*	2.2
	100	160.	20.	7.4	4.2	1.6*
	200	20.	1.8	1.2	1.0*	9.7
	AVAR†	80.	5.4	3.8	3.8	3.8
SP, M	50	140.	970.	40.	30.*	170.
	100	130.	1500.	20.*	40.	90.
	200	40.	270.	10.	7.0	6.5*
	AVAR†	90.	960.	6.7	6.7	6.7
WPI, M. NSA	50	150.	1300.	80.*	110.	300.
	100	30.	520.	40.*	60.	220.
	200	8.3*	90.	9.8*	10.	10.
	AVAR†	10.	210.	7.4	7.4	7.4

*Best estimator for this environment and sample size

† Asymptotic variance calculated for T = 200.

However, the other estimators are very close and for the larger sample sizes it appears that the optimum estimator is selected randomly. Thus, both ALS and the spectral estimators appear to have reached their asymptotic distribution at samples of 100.

Error process 2 is first order markov and therefore ALS is asymptotically efficient. However, it only captures 3 out of fifteen firsts. Apparently the three spectral estimators are approaching their asymptotic distribution just as rapidly as ALS, and from a comparison of the variances at 200, both are essentially there.

Error 3 is a strongly dependent process with large positive roots. Here ALS is a clear favorite as it dominates in 12 situations, in spite of the fact that the spectral estimators have roughly one third the asymptotic variance. Because of the steep decline in the spectrum of the errors at the very crucial low frequencies, we would expect HAN 3 to outperform the other two versions, and this in fact does happen. Nevertheless, these are all well above their asymptotic variances. It was in this case that Engle (6) showed that it would be possible for OLS to be better than ALS, if the variable was sufficiently concentrated at low frequencies. These results indicate that at least for the environments here, this does not happen.

Error 4 has a strong second order dependence but very little first order. Han 3 is the clear winner and appears to be only slightly above its asymptotic variance. ALS and OLS do behave very similarly and much worse than the spectral estimators.

Error 5 has an oscillation with a period of 8 time units. The spectral estimators take most of the prizes with a slight edge for HAN 3. ALS, as predicted by Engle, does much worse than OLS.

Overall, the results are very encouraging for the Hannan estimators. They are generally only slightly above their asymptotic variances over the range of observations examined, and even in simple situations do as well as the appropriate estimators. The biggest failure of the spectral estimators was in error 3 where the difficulty in estimating a steep slope of the

spectrum at low frequencies appeared to generate a large amount of extra noise in the estimator and enabled ALS to successfully compete at these sample sizes. HAN 3 was most able to model this behavior but the choice of n=2 may be in part responsible for this success.

A possible explanation for the failure of HAN 1 and HAN 2 to model the low frequency peak, is that the bandwidth chosen was too wide to accurately pick up the peak. Experiments with narrower bandwidths however, lead to no improvement.

The simple implications of these results, are that use of one of the Hannan estimators in place of OLS or ALS will not cost very much for these sample sizes but will possibly produce a much better estimate.

VI. Asymptotic Approximations

In order to speak more generally about these results we estimated a second order term in the asymptotic variance. In general the variance may be expressed as

$$\text{(16)} \qquad \text{var}\,(\hat{\beta}) = \frac{1}{T}\,\text{AVAR}\,(\hat{\beta})\,(1 + c/T)$$

so that for small samples the term c/T may be significant in explaining the variance. In this functional form, c has an easy interpretation; it is the number of observations at which the actual variance will be just twice the asymptotic variance (if c is negative a similar interpretation can be given). From using our observed values of variances, asymptotic variance and T we can fit c by running the regression

$$\text{(17)} \qquad \frac{T \cdot \text{VAR}(\hat{\beta})}{\text{A VAR}\,(\hat{\beta})} - 1 = c/T \quad .$$

TABLE VIII

COEFFICIENTS AND T-STATISTICS

(T*VAR/AVAR) -1 = c/T

ERROR	ESTIMATOR				
	OLS	ALS	HAN 1	HAN 2	HAN 3
1.	12.76 (1.72)	14.65 (1.88)	20.52 (2.53)	26.97 (2.88)	14.69 (1.91)
2.	8.57 (.97)	14.42 (1.59)	18.93 (2.08)	22.47 (1.99)	15.41 (1.66)
3.	-14.89 (-2.77)	-15.10 (-1.18	929.53 (5.11)	601.80 (4.15)	325.20 (4.85)
4.	11.57 (1.16)	10.32 (1.05)	109.92 (2.81)	91.40 (3.24)	48.72 (2.81)
5.	25.25 (.81)	-18.16 (-1.51)	368.70 (2.88)	143.99 (3.77)	268.23 (3.74)

TABLE IX

POOLED REGRESSIONS

COEFFICIENTS WITH T-STATISTICS

DEPENDENT VARIABLE: $\frac{T*VAR}{AVAR} - 1$

VARIABLE	OLS	ALS	HAN 1	HAN 2	HAN 3
1/T	8.65 (1.24)	1.23 (.25)	127.53 (2.43)	69.81 (2.05)	86.25 (3.41)
ERROR SLOPE/T			.003 (6.84)	.002 (6.97)	.00087 (4.21)

There is of course a serious question as to whether we can pool over x processes or error processes. We felt that pooling over x processes was not only legitimate but desirable since we would like to interpret our results in terms of a typical economic process.

We first estimated c for each estimator and error. These are presented in Table VIII along with the t-statistics. Notice that the values for c are generally not significantly different from zero for OLS and ALS and in fact are often negative suggesting that they do better than their asymptotic variances. The spectral estimators however, all have strongly positive and generally significant coefficients. Because the sizes differed so much, we decided that it was necessary to make an effort also to explain the differences between the errors.

In the light of the difficulties discussed in estimating the spectrum at low frequencies, we hypothesized that the steepness of the spectrum at low frequencies might be an important variable explaining how the spectral estimators behaved in different environments. The slope of the spectrum at zero frequency is zero, but at low frequencies it is proportional to the negative sine. The absolute value of the proportionality constant is

$$\text{SLOPE} = \left| \alpha_1/(1-\alpha_1)^2 + \alpha_2/(1-\alpha_2)^2 \right| \; / \; (1-\alpha_1)^2(1-\alpha_2)^2 \tag{18}$$

where α's are the roots of the error process.

For the spectral estimators the pooled equation was

$$\frac{T \cdot \text{VAR}(\hat{\beta})}{\text{A VAR}(\hat{\beta})} - 1 = c/T + c_1 \, \text{SLOPE}/T \tag{19}$$

Alternative forms used the first order serial correlation coefficient or a variety of functions of the asymptotic variances. This variable proved the best but a great deal of the variance remains unexplained and a finite sample theory would provide a better equation for this estimation.

Equation (17) was estimated for OLS and ALS over the entire set of 75 observations. The validity of the pooling assumption in each case was tested. Both OLS and ALS passed easily. Upon pooling, we found the coefficients are small, positive and insignificant.

The aggregation test was passed easily by HAN 1 and HAN 2 and only failed by a small margin for HAN 3. Therefore we pooled this data and estimated (19). The results in Table IX indicate that the slope variable is very significant and that the coefficient of reciprocal sample size is also significant although with a large standard error. Unless the slope is very great, HAN 2 and HAN 3 should be less than twice, and HAN 1 slightly more than twice the asymptotic variance at a sample size of 100. HAN2 appears to be the best on these grounds, reaching this point by T=70. However, in the presence of an ill-behaved error process with a steep low frequency spectrum, HAN 3 would be superior since the slope coefficient is less than half that for HAN 2. The estimator HAN 1 is dominated on both counts by the other two versions.

For none of these equations in either the original or pooled form could we reject at a 95% level, the hypothesis that the constant term is truely zero as indicated by the asymptotic theory. This provides some evidence for our estimating equations.

VII. Conclusions

We draw the following six conclusions from this study:

(1) We have formulated and programmed a computationally simple Hannan estimator which gives reasonable results. It is not iterative and is therefore not expensive to compute. Three versions which differ only for finite samples are examined.

(2) Based on the pooled results of five typical x-processes and five error processes, we find that our Hannan estimator will be approximately 2 times its asymptotic variance at sample size 100, provided the slope of the spectrum near 0 is not large. Under the traditional assumptions about error processes, the estimators reach this point for samples of size 20.

(3) On these same x- and error processes, we find no significant trend for ALS or OLS in terms of asymptotic variance.

(4) Thus, we expect that if the Hannan estimator's asymptotic variance is no more than half of that of OLS or ALS, then it will pay to switch to the Hannan estimator at sample size 100. This is the trade-off we observe between robustness and finite sample properties of the Hannan estimator.

(5) There are some principles of choice among the Hannan estimators. HAN 2 (piecewise constant disturbance spectrum) appears to dominate HAN 3 (autoregressive error process) as long as the slope of the error spectrum near zero is not large; otherwise this judgement is reversed. HAN 1 (rectangular moving average of residual periodogram) is dominated by the other two.

(6) Even at sample size 50, there appears to be no significant bias in any of the environments for any of the estimators .

BIBLIOGRAPHY

(1) Amemiya, Takeshi, "Generalized Least Squares with an Estimated Autocovariance Matrix", forthcoming in Econometrica.

(2) Box, George E.P., and Gwilym M. Jenkins, Time Series Analysis, Forecasting, and Control, Holden Day, 1970.

(3) Cooper, Phillip J., "Asymptotic Covariance Matrix of Procedures for Linear Regression in the Presence of First Order Serially Correlated Disturbances", Econometrica, (March 1972), 305-310.

(4) Duncan, D.B. and R. Jones, "Multiple Regression with Stationary Errors," American Statistical Association Journal, Vol 6 (December, 1966), 917-928.

(5) Engle, Robert F., "Band Spectrum Regression," International Economic Review, February 1974.

(6) Engle, Robert F., "Specification of the Disturbance for Efficient Estimation", Econometrica, forthcoming.

(7) Granger, C.W.J., "The Typical Spectral Shape of an Economic Variable," Econometrica, 34 (1966), p. 150.

(8) Griliches, Zvi and Rao, P., "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors", Journal of American Statistical Association, 64 (1969) 253.

(9) Hannan, E.J., Multiple Time Series, New York, John Wiley, 1970.

(10) Hannan, E.J., "Regression for Time Series," in Time Series Analysis, ed. by M. Rosenblatt, New York, John Wiley, 1963, 14-37.

(11) Knuth, Donald E., The Art of Computer Programming, Vol II - Seminumerical Algorithm, Addision Wesley, 1969.

(12) Maddala, G.S. "Generalized Least Squares with an Estimated Variance Covariance Matrix", Econometrica, January 1971.

(13) Parzen, E., "Multiple Time Series Modeling", in Multivariate Analysis, Academic Press, 1969.

(14) Watson, G.S. and E.J. Hannan, "Serial Correlation in Regression Analysis II," Biometrika 43 (1956), 436-448.