working paper
department
of economics

BAND SPECTRUM REGRESSIONS

by

ROBERT F. ENGLE

Number 96 November 1972

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

BAND SPECTRUM REGRESSIONS

by

ROBERT F. ENGLE

Number 96 November 1972

This is a preliminary draft --
not to be quoted without
the author's permission.

BAND SPECTRUM REGRESSIONS*

Ever since Hannan [11] first proposed regression analysis in the frequency domain, economists have been intrigued by the idea and have particularly explored its applications for distributed lag analysis, see Hannan [9],[10], Sims [13], Dhrymes [1], Fishman [5] and many others. Less attention has been lavished on static models and altogether, few useful empirical applications of spectral regressions have been found. The explanation for this centers on: (1) the need for large amounts of data, (2) somewhat complex computations involving judgmental parameters such as spectral windows and truncation points, and (3) rather minimal payoffs in terms of the gain in efficiency over comparable time domain techniques.

In this paper, I will argue against all three of these points by showing that frequency domain regressions have the standard small sample properties, that these do not require spectral windows, and computationally these procedures are easy to use, and most important, that they lead to very natural and simple solutions to such difficult problems as errors in variables and seasonality. In Section I, the theory of models using the full spectrum is restated in a slightly new fashion, Section II presents the comparable results for band spectral models with some tests of this specification, Section III describes the applications of these models to economic problems, and

---

*This research was partially supported by a grant from the Cambridge project. I am indebted to C. Sims, R. Hall, E. Kuh and E. Parzen for helpful comments.

IV describes simple computational techniques. Section V applies the method to a test of the permanent income hypothesis.

I. Full Spectrum Regressions

If we define the row vector w_k as

$$(1) \qquad w_k = \left(1, e^{i\theta_k}, e^{2i\theta_k}, \ldots, e^{(T-1)i\theta_k}\right)$$

where $\theta_k = 2\pi k/T$, then for a variable x_t; $t = 0, 1, \ldots, T-1$, $w_k x/\sqrt{T}$ is the k^{th} element of the finite Fourier transform of x where x is a column vector. The periodogram of x is defined as

$$(2) \qquad \hat{f}_x(\theta_k) = |w_k x|^2$$

and the cross periodogram between x and z is

$$(3) \qquad \hat{f}_{xz}(\theta_k) = (w_k x)^{\dagger}(w_k z)$$

where † means the complex conjugate of the transpose.*

The periodogram is well known (for example, see Jenkins and Watts [12]) to be a biased estimator of the spectrum. It is however asymptotically unbiased but inconsistent as the variance of each spectral estimator does not shrink as the sample becomes infinite. It is this inconsistency which forces the use of spectral windows or periodogram averaging to obtain spectral estimators. As we are concerned with the properties of our regression and only incidentally with spectral estimates we will not find it necessary to use these techniques.

---

*Here of course as $w_k x$ is only a scalar so the transpose is irrelevant.

Letting

(4) $$W = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{T-1} \end{pmatrix}$$

it is easily shown (Wahba [14]) that the normal trigonometric identities insure that the columns are orthonormal so that $W W^{\dagger} = I = W^{\dagger} W$ and W is a unitary matrix. Writing the vector $\tilde{x}$ as the Fourier transform of x times $\sqrt{T}$ we can transform a multivariate regression equation*

(5) $$y = x\beta + \epsilon$$

to obtain

(6) $$\tilde{y} = \tilde{x}\beta + \tilde{\epsilon}$$

Model (6) is a regression equation with complex random variables but fortunately this does not affect the standard results of classical regression models. The error properties of $\tilde{\epsilon}$ are given by

(7) $$\begin{aligned} \operatorname{var}(\tilde{\epsilon}) &= E\tilde{\epsilon}\tilde{\epsilon}^{\dagger} \\ &= E W \epsilon \epsilon^{\dagger} W^{\dagger} \\ &= W E \epsilon \epsilon^{\dagger} W^{\dagger} \\ &= \sigma^2 W \Omega W^{\dagger} \end{aligned}$$

and if $\Omega = I$ then $\tilde{\epsilon}$ is a complex spherical disturbance **vector**.

---

*This can of course be a distributed lag model but we assume that we have the same number of observations on all variables in contrast to some frequency domain analysis.

Assuming x (and therefore $\tilde{x}$) independent of ϵ, the Gauss-Markov theorem implies that the BLUE of (6) is

$$(8) \qquad \hat{\beta} = (\tilde{x}^{\dagger}\tilde{x})^{-1}\tilde{x}^{\dagger}\tilde{y}$$

with variance-covariance matrix

$$(9) \qquad \operatorname{var}\hat{\beta} = (\tilde{x}^{\dagger}\tilde{x})^{-1}\sigma^2 .$$

As there is only one BLUE of (5), the estimator (8) must be just OLS; and upon substituting for $\tilde{x}$ and $\tilde{y}$, we see that the unitary property of W guarantees this.

The estimator in (8) can be written in terms of the periodogram as

$$(10) \qquad \hat{\beta} = \left[\sum_{k=0}^{T-1} \hat{f}_{xx}(\theta_k)\right]^{-1} \sum_{k=0}^{T-1} \hat{f}_{xy}(\theta_k)$$

where $\hat{f}_{xx}(\theta)$ is a matrix of cross periodograms at each frequency and $\hat{f}_{xy}(\theta)$ is a vector of cross periodograms. Clearly, unsmoothed periodograms are required in estimating the parameters of the regression model. The estimator is consistent, not because each periodogram element approaches its spectral value, but because the sum of the elements approaches the sum of the spectral values which is just the total variance of the variable.

If in the original model (5) the disturbances were not spherical, then (7) must be examined more carefully. If end effects in the description of the disturbance are unimportant, either because Ω is defined as a circulant (Wabah [14]) or because T is large (Grenander and

Szego [7]) then $W \Omega W^{\dagger} = \text{diag}\, f_{\epsilon}(\theta)$ where the diagonal elements are the values of the spectrum of ε evaluated at the T harmonics. Therefore by constructing a matrix A with diagonal elements equal to $\sigma f_{\epsilon}^{-\frac{1}{2}}(\theta)$, we can multiply (6) by A and produce spherical disturbances. Thus (6), (8), (9) and (10) become respectively:

(11) $$A\tilde{y} = A\tilde{x}\beta + A\tilde{\epsilon}$$

(12) $$\hat{\beta} = (\tilde{x}^{\dagger}A^{\dagger}A\tilde{x})^{-1}(\tilde{x}^{\dagger}A^{\dagger}A\tilde{y})$$

(13) $$\text{var}(\hat{\beta}) = (\tilde{x}A^{\dagger}A\tilde{x})^{-1}\sigma^2$$

(14) $$\hat{\beta} = \left[\sum_{k=0}^{T-1} \hat{f}_{xx}(\theta_k)\, f_{\epsilon}^{-1}(\theta_k)\right]^{-1} \sum_{k=0}^{T-1} \hat{f}_{xy}(\theta_k) f_{\epsilon}^{-1}(\theta_k)$$

The problem of non-spherical disturbances in the time domain is seen to be merely a problem of heteroscedasticity in the frequency domain, and is therefore easier to handle in full generality.

To make this estimator operational, it is necessary to use an estimate of $f_{\epsilon}(\theta)$ from a first stage consistent estimate of the disturbances. It is generally assumed that this should be a consistent estimator of $f_{\epsilon}(\theta)$ at each frequency which requires the use of spectral windows, or equivalently, averaging over sections of the periodogram which are assumed to be smooth. The author knows of no careful examination of the properties of the various smoothing techniques for this special case although Duncan and Jones [2] suggest a one pass estimator which estimates the A's simultaneously by assuming the spectrum to be constant in sections and then using the estimated variances to correct for the heteroscedasticity.

II. Band Spectrum Regressions

For reasons which will be argued in Section III it may be useful to specify that a model applies for some but not all frequencies. Constructing a $T \times T$ matrix A with 1's on the diagonals corresponding to included frequencies and zero's elsewhere, a natural specification is

$$(15) \qquad A\tilde{y} = A\tilde{x}\beta + A\tilde{\epsilon}, \qquad E(A\tilde{\epsilon})(A\tilde{\epsilon})^{\dagger} = \sigma^2 A, \qquad \tilde{x} \text{ and } \tilde{\epsilon} \text{ independent.}$$

A is idempotent and symmetric, so (11)-(13) describe this estimator except that $A^{\dagger}A = A$. The periodogram version can be written more concisely by defining Σ' as the sum only over the included frequencies specified by A.

$$(16) \qquad \hat{\beta} = [\Sigma' \hat{f}_x(\theta_k)]\, \Sigma' \hat{f}_{x\,y}(\theta_k)$$

The estimator (12) will only be real if the matrix A has a symmetry axis which runs from northeast to southwest. That is, if frequency component k is included, then T-k must be included. Economically this merely means that both sines and cosines are to be included at each frequency. To establish this fact we define the $T \times T$ matrix C as:

$$(17) \qquad C = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdot 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \cdot 1 & 0 \\ \vdots & 0 & 1 \cdot & & \vdots \\ 0 & 1 & 0 \cdots & \cdots & 0 \end{pmatrix}$$

and employ the notation that a bar means the complex conjugate. We first observe the important symmetry property that $\bar{W} = C\,W$ and therefore

$\overline{\tilde{x}} = (\overline{Wx}) = \bar{W}x = CWx = C\tilde{x}$ as long as x is real. Since $CC = I$, $\tilde{x}^{\dagger} A \tilde{y} \equiv \tilde{x}^{\dagger} C C A C C \tilde{y} = (C\tilde{x})^{\dagger} C A C C \tilde{y} = \overline{\tilde{x}}^{\dagger} A \overline{\tilde{y}} = (\overline{\tilde{x}^{\dagger} A \tilde{y}})$ as long as $CAC = A$ which is the symmetry condition described. The other term of the estimator is real by the same argument.

The Gauss-Markov theorem assures that this estimator is best linear unbiased under the restrictions of A. However, if A is not a valid restriction then this estimator is inefficient relative to OLS. The more data points are excluded the more inefficient is the estimator and therefore there is a trade-off between efficiency and the possible bias due to misspecification.

A test of the exclusions is therefore necessary. Fortunately it is available in exactly the form always used in the time domain. The estimator of β is of course normal as long as y is normal. In addition, $\tilde{\epsilon}$ is normal (complex) as it is a linear combination of normal random variables. The residuals can be written as

$$(18) \qquad A\tilde{u} = A\tilde{y} - A\tilde{x}\hat{\beta} = (A - A\tilde{x}(\tilde{x}^{\dagger}A\tilde{x})^{-1}\tilde{x}^{\dagger}A)\tilde{y} = \tilde{M}\tilde{y} = \tilde{M}\tilde{\epsilon} .$$

$\tilde{M}$ is Hermitian $(\tilde{M}^{\dagger} = \tilde{M})$, idempotent and has rank equal to its trace. The trace of $\tilde{M}$ is $T' - K$ where T' is the number of included frequencies (the rank of A) and K is the number of regressors. An unbiased estimator of σ^2 is therefore

$$(19) \qquad s^2 = (A\tilde{u})^{\dagger} A\tilde{u}/T' - K$$

which is used in the standard tests of hypotheses and which is distributed as χ^2_{T-K} over its degrees of freedom.

A test of the exclusions A is obtained directly by following Fisher [4] in his exposition of the classic Chow tests. Replacing the word Hermitian for symmetric, unitary for orthogonal, and $\dagger$ for $'$ the argument is unaffected. Thus, letting $\tilde{u}^*$ be the unrestricted frequency domain residuals obtained in this case by using only some frequencies ($\tilde{u}^* \equiv A\tilde{u}$), and $\tilde{u}$ be the restricted residuals from forcing the β's to fit all frequencies, the statistic

$$(20) \qquad F = \frac{(\tilde{u}^\dagger \tilde{u} - \tilde{u}^{*\dagger}\tilde{u}^*) / (\mathrm{tr}\,\tilde{M} - \mathrm{tr}\,\tilde{M}^*)}{\tilde{u}^{*\dagger}\tilde{u}^*/\mathrm{tr}\,\tilde{M}^*}$$

is distributed as F with $\mathrm{tr}\,\tilde{M} - \mathrm{tr}\,\tilde{M}^*$ and $\mathrm{tr}\,\tilde{M}^*$ degrees of freedom under the null hypothesis.

III. Applications

In the time domain it is very common to exclude some periods such as wars or strikes because they do not conform to the model. Sometimes these exclusions are tested and sometimes tests for broad structural shifts over time are conducted. However, there is little discussion of whether the same model applies to all frequencies. It may be too much to ask of a model that it explain both slow and rapid shifts in the variables, or both seasonal and non-seasonal behavior. It is at least reasonable to test the hypothesis that the same model applies at various frequencies.

Furthermore, since the typical spectral shape of economic variables (Granger [6]) has a strong peak at low frequencies, these periodogram components completely dominate the parameter estimates. Often the sum of only the first three or four periodogram elements gives the same estimate as does the full OLS. This is merely an observation that the fitted line will pass exactly through an outlying value of x. While this experimental design gives relatively confident parameter estimates, we have little information about the fit if in fact the specification is not valid for the outlying x. Thus, it may be very useful to test the inclusion of the long term fluctuations in the model.

The relatively large literature on seasonal adjustment suggests that there may be no unique best way to seasonally adjust data; some procedures extract too much, and others too little, and in general there is much concern for what is done to the other components of the series. In a careful discussion of seasonal adjustment, Grether and Nerlove [8] point out that

> "the idea that an economic time series may be divided meaningfully into several unobserved components appears to have been firmly established in Economics since the time of Jevons. In itself, the division...is of little significance; it is, rather, that the components are themselves ascribable to separate and distinct groups of causes or influences."

It may be reasonable to estimate a different model for seasonal and non-seasonal components, rather than to attempt to extract exactly the seasonal portion. At least the exclusion of all seasonal components from the model presents a method for avoiding the question of seasonality.

As an example, suppose we have monthly data so that the seasonal frequencies and their harmonics have periodicities of 12, 6, 4, 3, 2.4, and 2 months. If it is assumed that the seasonality has been extremely

regular over the sample period, it is enough to eliminate these six data points and their 5 symmetrical counterparts of the A matrix.*
However, if the seasonal factors are assumed to be slowly changing then a narrow band around each harmonic should be extracted. The more flexible the desired version of seasonality, the wider should be the excluded bands. This approach to seasonality has the additional merit that as more flexible seasonal assumptions are adopted, the final degrees of freedom of the regression are correspondingly decreased.

A second class of models for which band spectral regressions are an appropriate estimation technique, are errors-in-variables models. In general, we need additional information to form consistent estimates of these models, whether it be variance information or properly specified instrumental variables. Very often it is reasonable to describe the additional information in terms of the frequency decomposition of the variance of the measurement error. If the error component is assumed to be confined to a particular frequency band, then a natural procedure is to eliminate that frequency band from the regression. This is essentially the technique used by Friedman in defining permanent income as a moving average of measured income. This filter eliminated high frequency noise which he calls the transient component of income.

To be specific, suppose

(21) $$y = x\beta + \epsilon ; \qquad z = x + v ,$$

---

*As Chris Sims has pointed out to me, this is exactly comparable to using 11 seasonal dummies in the time domain.

where x is unobservable but z can be measured. Assuming that z is independent of ϵ and that v has no variance at some frequencies (notice that we do not need to assume that v is independent of x) we can construct the matrix A which includes only frequencies where the variance of v is zero. The bias of the estimator corresponding to (12) is

$$E(\hat{\beta} - \beta) = E(\tilde{z}^{\dagger} A \tilde{z})^{-1} \tilde{z}^{\dagger} A \tilde{v} \beta \ .$$

If the expected value of the square of $\tilde{v}$ is zero then $\tilde{v}$ must be identically zero at that frequency and therefore the bias is zero. This implies that v is a process with a deterministic component. To avoid this assumption, consider a sequence of models where z has an ever increasing signal to noise ratio in the included frequencies. The upper bound on the bias depends on this ratio and tends toward zero in the limit. This argument provides a justification for eliminating some frequencies even though the variance of v is not assumed identically zero at the included frequencies on the grounds that the signal of these frequencies is much stronger than the noise component.

IV. Computational Considerations

The procedures described in the previous sections are direct analogues of OLS where Fourier transforms replace the real data. Most regression programs do not allow complex data; however, by simple preprocessing of the real data, the same results can be obtained.

Defining

(22) $$x^{*} = W^{\dagger} A W x$$

which is a real T element data vector obtained by taking the inverse Fourier transform of the Fourier transform times A, we can rewrite (11)-(13) as:

$$(23)\qquad y^* = x^*\beta + \epsilon^*$$

$$(24)\qquad \hat{\beta} = (x^{*\prime}x^*)^{-1} x^{*\prime} y^*$$

$$(25)\qquad \mathrm{var}(\hat{\beta}) = (x^{*\prime}x^*)^{-1}\sigma^2$$

The conventional regression packages will therefore provide band spectral or generalized least squares regression output if the data is transformed by (22).

The sampling statistics are however slightly incorrect if A is not of full rank. $E\,u^{*\prime}u^* = \sigma^2(T' - K)$ while the regression program will use T instead of T'. The use of this type of seasonal adjustment must be accompanied by a decrease in the degrees of freedom.

In the errors-in-variables models however, a slightly different procedure may be desired. If eq. (21) can be written as

$$(26)\qquad y = x^*\beta + \epsilon$$

OLS on this model gives the same estimate as (24) since

$$\begin{aligned} x^{*\dagger}y^* &= x^\dagger W^\dagger A W W^\dagger A^\dagger W y \\ &= x^\dagger W^\dagger A A^\dagger W y \\ &= x^{*\dagger} y \end{aligned}$$

and A is idempotent and symmetric. The appropriate error measure is now $u'u/T - K$ as an estimate of σ^2 and thus the package output is correct as given. If (26) is not appropriate, then the estimate of σ^2 will alone be affected and will be an overestimate. Filtering only the exogenous variables in order to remove the effects of errors at some frequencies does not alter the statistical properties of the conventional regression output except that included in the standard error is a component of the measurement error.

A variation on the filtering procedure will therefore not affect the validity of the regression result. A useful alternative form of filtering when only a small band is to be left in the variable is the band filtering procedure suggested by Granger and Hatanaka [6] which consists of first demodulating the series by multiplying by a sine and cosine wave of constant frequency. These results are then filtered with a low pass filter and finally the two series are remodulated by multiplying by the same sine and cosine series. This corresponds to the inclusion of a band of frequencies using weights which are not necessarily zero or one. In some respects this filter is easier to apply than the ones described above and has given useful results in other studies [2].

V. Another Test of the Permanent Income Hypothesis

Friedman hypothesized that permanent consumption was a linear function of permanent income. However, the relation between transitory income and consumption was not as clear and, in fact, the propensity to consume out of transitory income is often assumed to be zero. Noticing that transitory components of a series are primarily high frequency components, the Friedman hypothesis suggests that regressions using only high frequency components would behave differently from those with only low frequencies. In particular, the marginal propensity to consume would be substantially lower.

In a **simple** test of the hypothesis that permanent and transitory components have the same MPC, quarterly data on money income and consumption from the first quarter of 1946 through the third quarter of 1971 were examined using the techniques described above. The transient component was assumed to be that part with frequencies higher than two years per cycle. This meant that one quarter of the data points were in the portion of the spectrum generating the permanent series and the rest generated the transient series. By variance, however, the low frequencies are by far the most important in both of these series.

The results are given in the table. The original series were set to have mean zero but a constant was included to adjust for round off errors. It was always very small and is not reported; similarly no degree of freedom is extracted for its estimation.

Table 1

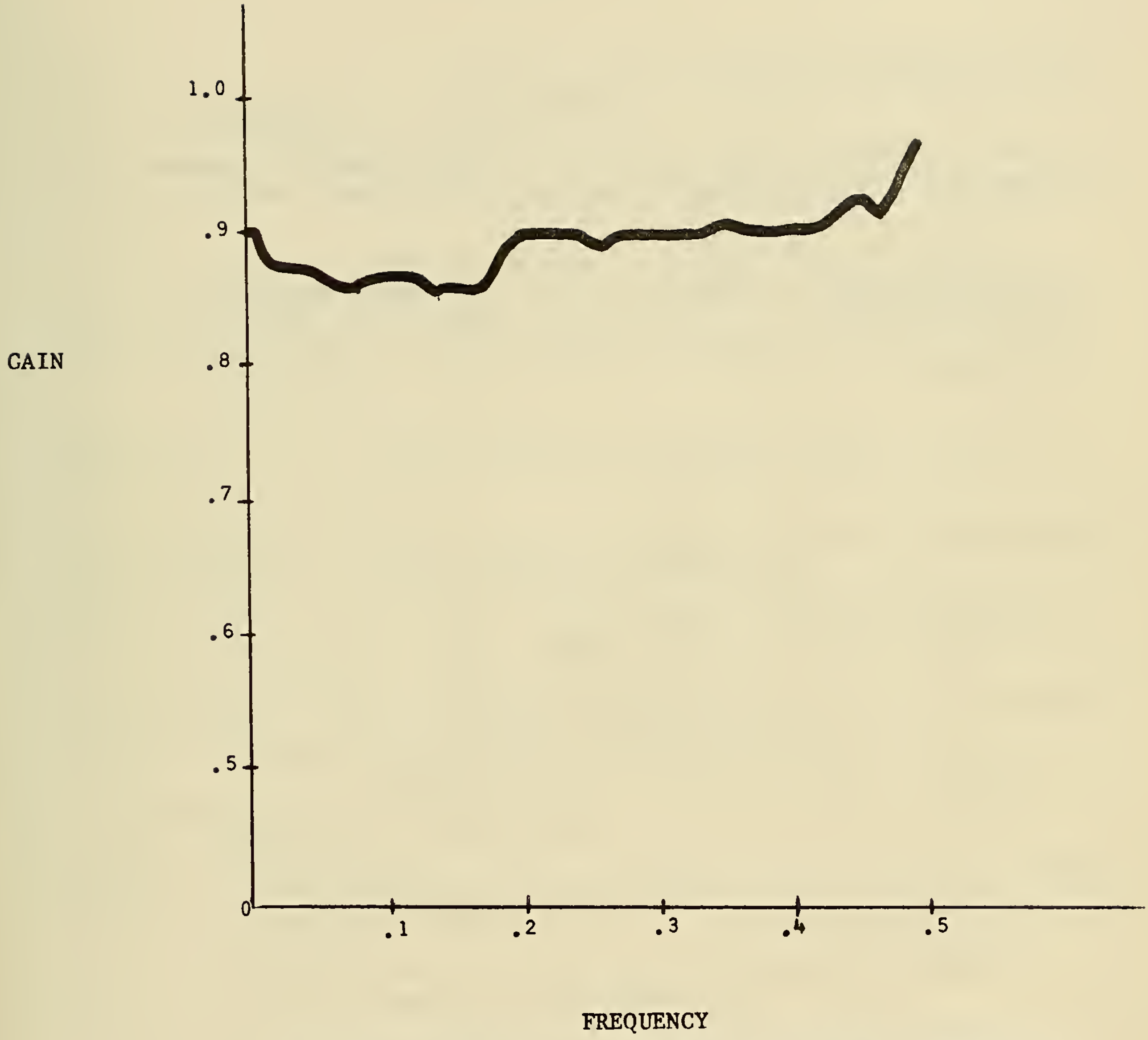
Estimates of the Marginal Propensity to Consume

	M.P.C	St. Error	T-Stat	D.F.	SSR
All	.89541	.00293	305.4	101	1815.670
Transitory (high frequency)	.89397	.01056391	84.6	75.5	366.290
Permanent (low frequency)	.89545	.01095428	81.7	24.5	1449.230

$F_{1,100}$ = .0065 for testing different specification. Forming the F statistic from (20) we get the extraordinarily low value of .0065.

Clearly, it is not possible to reject at any reasonable level the hypothesis that the propensity to consume differs between permanent and transitory series. The sum of squared residuals is remarkably similar in the pooled and unpooled case.

To examine whether the particular choice of a dividing point was responsible for the startling result, the gain or transfer function from income to consumption was calculated. This indicates the static regression coefficient, frequency by frequency, as long as the series are in phase. It can be interpreted as the long run multiplie at that frequency. This is shown in Figure 1. While there is some variation with frequency, it is not substantial and in general this very simple consumption function exhibits a great deal of stability. In particular, it does not appear to be true that higher frequencies have a smaller or even more erratic gain.



Marginal propensity to consume at different frequencies.

FIGURE 1

BIBLIOGRAPHY

[1] Dhrymes, Phoebus J. Distributed Lags: Problems of Estimation and Formulation, Holden-Day, 1971.

[2] Duncan, D.B. and R. Jones. "Multiple Regression with Stationary Errors," American Statistical Association Journal, December 1966.

[3] Engle, Robert F. and Duncan Foley. "A Supply Function Model of Aggregate Investment," M.I.T. Department of Economics Working Paper No. 89 , August 1972.

[4] Fisher, Franklin M. "Test of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," Econometrica, March 1970.

[5] Fishman, G.S. Spectral Methods in Economics, Harvard University Press, 1968.

[6] Granger, C.W.J. and M. Hatanaka. Spectral Analysis of Economic Time Series, Princeton University Press, 1964.

[7] Grenander, U. and Szego. Toeplitz Forms and Their Applications, University of California Press, 1968.

[8] Grether, D.M. and M. Nerlove. "Some Properties of Optimal Seasonal Adjustment," Econometrica, September 1970.

[9] Hannan, E.J. "The estimation of a lagged regression relation," Biometrika, Volume 54, 1967.

[10] _____, "The Estimation of Relationships Involving Distributed Lags, Econometrica, January 1965.

[11] _____, "Regression for Time Series," in Time Series Analysis ed. by M. Rosenblatt, John Wiley, 1963.

[12] Jenkins, G. and Donald G. Watts. Spectral Analysis and Its Application, Holden-Day, 1968.

[13] Sims, Christopher A. "Are There Exogenous Variables in Short-Run Production Relations," Discussion Paper No. 10, Center for Economic Research, University of Minnesota, August 1971.

[14] Wahba, Grace. "On the Distribution of Some Statistics Useful in the Analysis of Jointly Stationary Time Series," The Annals of Mathematical Statistics, Vol. 39, no. 6.