Volume: 10 Number: 01 Page: 95-111 ISSN No: 1303-6319 Cilt: 10 Sayı: 01

ESTIMATION AND TESTING FOR COINTEGRATION: A SPECTRAL REGRESSION APPROACH

Yılmaz AKDİ* **David A. DICKEY****

ABSTRACT

A popular topic in the econometrics and time series area is the cointegrating relationship among the components of a vector autoregressive time series. The problem became important after the work of Engle and Granger (1987) and has been addressed by many authors: Johansen (1988), Stock and Watson among many others. Engle and Granger's least squares method and Johansen's conditional maximum likelihood method have received the most attention. These tests are routinely applied to economic time series because the notion of cointegration has a natural interpretation. Our method uses low frequency components of the cross periodogram to estimate the cointegration relationship between cointegrated time series. The method improves the results of ordinary least squares method proposed by Engle and Granger in some cases.

Keywords: Time series, Cointegration, Periodogram ordinate, Spectral regression.

1. INTRODUCTION

Unit root tests comprise a standard diagnostic tool in applied time series analysis. There are several procedures to test for a unit root (e.g. Dickey and Fuller, 1979). Test procedures have also been developed to test for seasonal unit roots (Dickey, Hasza and Fuller (1984), Hylleberg, Engle, Granger and Yoo, 1990). Dickey and Pantula (1987) propose a procedure to test for multiple unit roots. Series with unit roots described as integrated. Akdi and Dickey (1998) developed a procedure to test for a unit root using the periodogram ordinates of a univariate time series.

Time series variables with a common (joint) stochastic trend form a cointegrated system. That is, if all of the individual time series are integrated, say of order one, it is sometimes possible that some linear combination of the series will be integrated of order zero (that is, stationary). Thus, the multiple time series $\underline{Y}_t$ is a vector of nonstationary time series, but there exists a vector (or a matrix) $\underline{\beta}$ such that $\underline{\beta}'\underline{Y}_t$ is a stationary system. This notion is known as cointegration and $\underline{\beta}$ is called the cointegrating vector (or matrix).

Engle and Granger (1987) have proposed an estimation procedure for the cointegrating vector. They used a regression approach to estimate the cointegrating vector $\underline{\beta}$. Johansen (1988) gave an estimation procedure that has become very popular. Levy (2002) take adventage of a squared coherency, phase and gain to study the cointegrating relationship for a bivariate cointegrated system. He derives some restrictions by

*Prof. Dr., Department of Statistics, Faculty of Science, Ankara University, Ankara, e-mail: akdi@science.ankara.edu.tr

**Department of Statistics, Nort Carolina State Universty, Raleigh, USA

studying cross-spectral properties of a cointegrated bivariate system. Boswijk and Lucas (2003) considers a semi-nonparametric cointegration test by using LM-testing principles. Breitung (2002) variance ratio testing procedure to test for a unit root and he suggests a generalization of the variance ratio test for cointegration. Chen and Hurvich (2003) study the asymptotic distribution of a tapered narrow-band least squares estimator of the cointegrating vector $\underline{\beta}$ in the framework of fractional cointegration. Deo and Hurvich (2001) study the estimators based on the log periodogram regression and they obtain the asymptotic bias and variance. They suggest to use low frequencies in the context of the long memory stochastic volatility model. Finite sample properties of spectral regression estimators have been studied by Chambers (2001) with simulation. Marunicci (2000) deals with a somewhat related problem. He considers spectral regression for cointegrated time series with long memory innovations. He provides a functional central limit theorem as a quadratic forms in nonstationary fractionally integrated processes. We investigate an estimation procedure for the cointegrating vector based on the periodogram. For simplicity, bivariate series are considered in detail and the extensions to higher dimensional autoregressive processes are discussed.

In section 2, some notation and definitions are introduced. Section 3 deals with the estimation procedures and consistency results. Several estimation strategies are given. These are regression using the frequency components of the series and different number of frequencies used. Simulation results indicate that using all frequencies in the regression gives the worst result. Even in this case, better results are obtained than those of least squares. Section 4 discusses the extension of the method to higher dimensional processes and finally Section 5 include real data example and a Monte-Carlo simulation study.

2. NOTATION AND MOTIVATION

Consider a first order vector autoregressive (VAR(1)) time series model

$$\underline{Y}_t = A\underline{Y}_{t-1} + \underline{e}_t \tag{1}$$

where $\underline{e}_t$ is a sequence of independent normally distributed random variables with mean vector $\underline{0}$ and variance covariance matrix V. Note that the process is stationary if all eigenvalues of A are less than 1 in absolute value and nonstationary otherwise.

If the coefficient matrix A has distinct eigenvalues, then there are matrices Q and M such that $AQ = QM$ where M is a diagonal matrix of the eigenvalues of A. The transformation $\underline{Z}_t = Q^{-1}\underline{Y}_t$ gives the canonical form of the series. For example, a bivariate series with

$$A = \begin{bmatrix} 1.8 & -0.8 \\ 1.2 & -0.2 \end{bmatrix}, \qquad M = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}, \qquad Q = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$$

the canonical form of the series can be obtained by setting $A = QMQ^{-1}$ as $\underline{Z}_t = M\underline{Z}_{t-1} + \underline{\xi}_t$ where $\underline{Z}_t = (U_t, S_t)'$ and $U_t = U_{t-1} + \xi_{1,t}$ and $S_t = 0.6S_{t-1} + \xi_{2,t}$. Substacting $\underline{Y}_{t-1}$ from both sides of (1) we have $\underline{Y}_t - \underline{Y}_{t-1} = (A - I)\underline{Y}_{t-1} + \underline{e}_t = \Pi\underline{Y}_{t-1} + \underline{e}_t$. That is, $\nabla\underline{Y}_t = \Pi\underline{Y}_{t-1} + \underline{e}_t$. In our example, notice that the matrix Π can be written as

$$\Pi = A - I = \begin{bmatrix} 1.8 & -0.8 \\ 1.2 & -0.2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.8 & -0.8 \\ 1.2 & -1.2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 1.2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \underline{\alpha\beta'}$$

such that $W_t = 0.6W_{t-1} + \xi_t$ is stationary. Here, $W_t = \underline{\beta}'\underline{Y}_t$.

The inverse transformation displays the cointegration relationship. One way to estimate this cointegrating vector in a bivariate case is to regress $Y_{2,t}$ on $Y_{1,t}$. If the residual series is stationary, then the bivariate series is cointegrated. This has been studied by Engle and Granger (1987) and consistency properties have been discussed by Stock (1987).

In a dimension p process, $\underline{Y}_t = A\underline{Y}_{t-1} + \underline{e}_t$ subtract $\underline{Y}_{t-1}$ from both sides to obtain

$$\underline{Y}_t - \underline{Y}_{t-1} = (A - I)\underline{Y}_{t-1} + \underline{e}_t = \Pi\underline{Y}_{t-1} + \underline{e}_t$$

where $\Pi = A - I = \underline{\alpha}\,\underline{\beta}'$. If Π.is of rank r, $(0 < r < p)$ and $\underline{\alpha}'\underline{\beta}$. is nonsingular , the series is a linear combination of r stationary and $p - r$ unit root canonical series, as used by Johansen (1988). We have shown Π for our bivariate example.

When a bivariate series $\underline{Y}_t$ is given, each component of the series can be considered as a sum of a stationary and a nonstationary series:

$$\begin{aligned} Y_{1,t} &= q_{11}U_t + q_{12}S_t \\ Y_{2,t} &= q_{21}U_t + q_{22}S_t \end{aligned} \qquad (2)$$

where U_t and S_t represent nonstationary (unit root series) and stationary series, respectively. From this representation, it can be seen that $Y_{2,t} - (q_{21}/q_{11})Y_{1,t} = (q_{22} - (q_{21}q_{12}/q_{11}))S_t = cS_t$ is stationary. That is, $\underline{\beta}'\underline{Y}_t$ is stationary when $\underline{\beta}' = (-q_{21}/q_{11}, 1)$. That is, $\underline{\beta}'$ is a cointegrating vector. Of course, the coefficient matrix A is unknown and thus has to be estimated.

Periodograms: Assume that a bivariate series $\underline{Y}_t$ with components $Y_{1,t}$ and $Y_{2,t}$ is given and assume that the representation in (2) is available. For any univariate time series $X_t, t = 1,2,3,\ldots,n$, the periodogram ordinate at the frequency w_k is defined by

$$I_X(w_k) = \frac{n}{2}\left(a_k^2 + b_k^2\right)$$

where $a_k = \frac{2}{n}\sum_{t=1}^{n}(X_t - \mu)\cos(w_k t)$, $b_k = \frac{2}{n}\sum_{t=1}^{n}(X_t - \mu)\sin(w_k t)$. Here, μ is the mean of the series and when $w_k = 2\pi k/n$, since $\sum_{t=1}^{n}\cos(w_k t) = \sum_{t=1}^{n}\sin(w_k t) = 0$ the periodogram is invariant to the mean whether it is known or estimated.

Given a set of observations $\{\underline{Y}_1, \underline{Y}_2, ..., \underline{Y}_n\}$ the periodograms of U_t and S_t are defined as in (3) below:

$$I_{u,n}(w_k) = \frac{n}{2}\left(a_{u,k}^2 + b_{u,k}^2\right),\ I_{s,n}(w_k) = \frac{n}{2}\left(a_{s,k}^2 + b_{s,k}^2\right) \tag{3}$$

and the real part of the cross periodogram ordinate between U_t and S_t is defined as

$$\mathrm{Re}\,al\{I_{u,s,n}(w_k)\} = \frac{n}{2}\left(a_{u,k}a_{s,k} + b_{u,k}b_{s,k}\right) \tag{4}$$

where $w_k = 2\pi k/n, k = 0,1,2,...,[n/2]$ (here, $[x]$ denotes the integer part of x). The Fourier coefficients for the series U_t and S_t are

$$\begin{aligned} a_{u,k} &= \frac{2}{n}\sum_{t=1}^{n}U_t\cos(w_k t) \ , \quad a_{s,k} = \frac{2}{n}\sum_{t=1}^{n}S_t\cos(w_k t). \\ b_{u,k} &= \frac{2}{n}\sum_{t=1}^{n}U_t\sin(w_k t) \ , \quad b_{s,k} = \frac{2}{n}\sum_{t=1}^{n}S_t\sin(w_k t). \end{aligned} \tag{5}$$

For any stationary time series X_t, the normalized periodogram ordinate is asymptotically distributed as chi-square with two degrees of freedom.

3. ESTIMATION AND CONSISTENCY RESULTS

In this section, an estimation procedure for the cointegrating vector based on the periodogram ordinates is discussed. Note that, if a vector process is stationary, then each component is marginally stationary. Akdi and Dickey (1998) show that the periodogram of the unit root process U_t satisfies

$$\frac{4k^2\pi^2}{n^2}I_{u,n}(w_k) \xrightarrow{D} \sigma^2\left(Z_1^2 + Z_2^2\right) \ as\ n \to \infty \tag{6}$$

Here, Z_1 and Z_2 are independent standard normal random variables. This result can be used to test for a unit root. The critical values of the distribution under the null hypothesis of a unit root is tabulated in Akdi and Dickey (1998). For a stationary time series $\{S_t, t = 1,2,3,...,n\}$, $I_{s,n}(w_k) = (n/2)\left(a_k^2 + b_k^2\right) = O_P(1)$ as shown in Fuller (1996). This means that the unit root dominates all the asymptotic properties. The cited references show that $a_{u,k} = O_P(n^{1/2})$, $a_{s,k} = O_P(n^{-1/2})$ and $b_{u,k} = O_P(n^{1/2})$, $b_{s,k} = O_P(n^{-1/2})$. Thus we have

$$\begin{aligned} \mathrm{Re}\,al\{I_{u,s,n}(w_k)\} &= \frac{n}{2}\left(a_{u,k}a_{s,k} + b_{u,k}b_{s,k}\right) \\ &= O(n)\left[O_P(n^{1/2})O_P(n^{-1/2}) + O_P(n^{1/2})O_P(n^{-1/2})\right] = O_P(n). \end{aligned}$$

Theorem 1. Consider the transformation in (2). For ease of notation and to emphasize the analogy to regression, write X_t for $Y_{1,t}$ and Y_t for $Y_{2,t}$. For each fixed k, the ratio

Volume: 10 Number: 01 Page: 95-111 ISSN No: 1303-6319 Cilt: 10 Sayı: 01

$y_k = \mathrm{Real}\{I_{XY}(w_k)\}$ to $x_k = I_X(w_k)$ is a consistent estimator of the ratio q_{21}/q_{11}. That is, for each fixed k

$$C_n(w_k) = \frac{\mathrm{Real}\{I_{XY}(w_k)\}}{I_X(w_k)} \xrightarrow{P} \frac{q_{21}}{q_{11}}, \quad as \ n \to \infty .$$

Proof: Notice that

$$b_{X,k}(w_k) = \frac{2}{n}\sum_{t=1}^{n} X_t \sin(w_k t) = \frac{2}{n}\sum_{t=1}^{n}(q_{11}U_t + q_{12}S_t)\sin(w_k t) = q_{11}b_{u,k} + q_{12}b_{s,k}$$

and similarly $a_{X,k}(w_k) = q_{11}a_{u,k} + q_{12}a_{s,k}$ and thus

$$n^{-2}I_X(w_k) = (2n)^{-1}\left(a_{X,k}^2 + b_{X,k}^2\right) = (2n)^{-1}\left[\left(q_{11}a_{u,k} + q_{12}a_{s,k}\right)^2 + \left(q_{11}b_{u,k} + b_{12}a_{s,k}\right)^2\right]$$
$$= (2n)^{-1}\left[q_{11}^2\left(a_{u,k}^2 + b_{u,k}^2\right)\right] + O_P(n^{-1}) = q_{11}^2\left(n^{-2}I_u(w_k)\right) + O_P(n^{-1}).$$

Likewise the cross periodogram ordinate of X_t and Y_t can be written as follows:

$$\mathrm{Real}\left\{n^{-2}\, I_{XY}(w_k)\right\} = q_{11}q_{21}\left(n^{-2}I_u(w_k)\right) + O_P(n^{-1})$$

and thus,

$$C_n(w_k) = \frac{\mathrm{Real}\{I_{XY}(w_k)\}}{I_X(w_k)} = \frac{\mathrm{Real}\left\{n^{-2}I_{XY}(w_k)\right\}}{n^{-2}I_X(w_k)} = \frac{q_{11}q_{21}}{q_{11}^2} + O_P(n^{-1})$$

which completes the proof.

As a result of Theorem 1, when we fix the number of frequencies in the regression, we still have the concictency. That is, consider the following regression model

$$y_k = \beta x_k + \xi_k, \ k = 1,2,3,\ldots,m \tag{7}$$

where y_k is the real part of the cross periodogram ordinate of the series X_t and Y_t as defined in (4) at frequency k, and x_k is the periodogram ordinate of the series X_t. Then the ordinary least squares estimate $\hat{\beta}_{1,n}$ of β in model (7) is a consistent estimator of the ratio q_{21}/q_{11}. That is,

$$\hat{\beta}_{1,n} = \frac{\sum_{k=1}^{m} x_k y_k}{\sum_{k=1}^{m} x_k^2} = \frac{\frac{1}{m}\sum_{k=1}^{m} x_k y_k}{\frac{1}{m}\sum_{k=1}^{m} x_k^2} \xrightarrow{P} \frac{q_{21}}{q_{11}} \quad as \ n \to \infty \tag{8}$$

Moreover, using the intercept term in the regression,

$$y_k = \alpha + \beta x_k + \xi_k, \ k = 1,2,3,\ldots,m \tag{9}$$

the ordinary least squares estimate of β is still a consistent estimator of the ratio q_{21}/q_{11}. That is,

$$\hat{\beta}_{1,n} = \frac{\sum_{k=1}^{m}(x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^{m}(x_k - \bar{x})^2} \xrightarrow{P} \frac{q_{21}}{q_{11}} \quad as \ n \to \infty . \tag{10}$$

4. HIGHER ORDER SERIES

In this section, higher order and higher dimension vector autoregressive time series are considered. Consider the following time series model

$$\underline{Y}_t = A_1\underline{Y}_{t-1} + A_2\underline{Y}_{t-2} + \ldots + A_p\underline{Y}_{t-p} + \underline{e}_t \qquad (11)$$

where $\underline{Y}_t$ is a k-variate random vector, A_i's are appropriate matrices and $\underline{e}_t$ is a sequence of $i.i.d.$ random vectors with mean-vector $\underline{0}$ and variance-covariance matrix Σ. Subtracting $\underline{Y}_{t-1}$ from both sides, the model becomes $\nabla\underline{Y}_t = \Pi\underline{Y}_{t-1} + B_1\nabla\underline{Y}_{t-1} + \ldots + B_p\nabla\underline{Y}_{t-p} + \underline{e}_t$ where $\Pi = I - A_1 - A_2 - \ldots - A_p$, $B_i = -(A_{i+1} + A_{i+2} + \ldots + A_p)$. The number, $p-r$, of unit roots of the characteristic equation $|\, m^p I - m^{p-1}A_1 - \ldots - A_p \,| = 0$ is the rank deficiency of Π. In this case, there exist rank r matrices α and β such that $\Pi = \alpha\beta'$. We assume that each element of the response vector $\underline{Y}_t$ has a unit root and stationary first difference. That is, $\underline{Y}_t$ is integrated of order 1. Thus, it is clear that

$$\nabla\underline{Y}_t = A_1\underline{Y}_{t-1} + B\underline{W}_t \qquad (12)$$

where $\underline{W}_t$, being a linear combination of stationary series, is stationary. Thus, the problem reduces to estimating the coefficient matrix Π.

Using a similar argument to that in section 2 and partitioning $\underline{Y}_t$, the following representation is available for the higher dimensional processes, too.

$$\begin{aligned} \underline{Y}_{1,t} &= Q_{11}\underline{U}_t + Q_{12}\underline{S}_t \\ \underline{Y}_{2,t} &= Q_{21}\underline{U}_t + Q_{22}\underline{S}_t \end{aligned} \qquad (13)$$

where $\underline{U}_t$ represents the components with unit roots and $\underline{S}_t$ represents stationary components of the series. Therefore, the problem reduces to estimating $Q_{21}Q_{11}^{-1}$. We are assuming Q_{11} is $(p-r)\times(p-r)$ and of full rank, by which assumption we have identified as a set of series to construct $\underline{Y}_{1,t}$ involves all of the $p-r$ nonstationary comment trends. That is, $Q_{11}^{-1}\underline{Y}_{1,t}$ is $\underline{U}_t$ plus stationary components. Only in this way, $\underline{Y}_{1,t}$ can be used to remove all the nonstationary components from the other series. Analogous to the bivariate case $Q_{21}Q_{11}^{-1}$ is to be estimated. Exactly as in the bivariate case, the order of the unit root parts of the Fourier transforms of the data dominate. Specifically, the Fourier coefficient matrices for $\underline{Y}_t$ are:

$$\underline{A}_{1,k} = \frac{2}{n}\sum_{t=1}^{n}\underline{Y}_{1,t}\cos(w_k t) \;,\; \underline{B}_{1,k} = \frac{2}{n}\sum_{t=1}^{n}\underline{Y}_{1,t}\sin(w_k t)$$

$$\underline{A}_{2,k} = \frac{2}{n}\sum_{t=1}^{n}\underline{Y}_{2,t}\cos(w_k t) \;,\; \underline{B}_{2,k} = \frac{2}{n}\sum_{t=1}^{n}\underline{Y}_{2,t}\sin(w_k t)$$

and those for the series $\underline{U}_t$ and $\underline{S}_t$ are similar and denoted by lower case letters correspondingly; $\underline{a}_{u,k}$, $\underline{a}_{s,k}$, $\underline{b}_{u,k}$, $\underline{b}_{s,k}$. The relationship between the Fourier coefficients

$\underline{A}_{i,k}$, $\underline{B}_{i,k}$, $i=1,2$ and $\underline{a}_{u,k}$, $\underline{a}_{s,k}$, $\underline{b}_{u,k}$, $\underline{b}_{s,k}$ is as follows:

$$\underline{A}_{1,k} = Q_{11}\underline{a}_{u,k} + Q_{12}\underline{a}_{s,k} \quad , \quad \underline{B}_{1,k} = Q_{11}\underline{b}_{u,k} + Q_{12}\underline{b}_{s,k}$$
$$\underline{A}_{2,k} = Q_{21}\underline{a}_{u,k} + Q_{22}\underline{a}_{s,k} \quad , \quad \underline{B}_{2,k} = Q_{21}\underline{b}_{u,k} + Q_{22}\underline{b}_{s,k}$$

The periodogram ordinate of $\underline{Y}_{1,t}$ can be calculated as follows:

$$P_{11}(w_k) = \frac{n}{2}\left(A_{1,k}\,A_{1,k}{}' + B_{1,k}\,B_{1,k}{}'\right)$$

and the real part of the cross periodogram ordinate will be

$$R_{12}(w_k) = \frac{n}{2}\left(A_{1,k}\,A_{2,k}{}' + B_{1,k}\,B_{2,k}{}'\right).$$

Now define the *cointegrating spectrum* as

$$C_n(w_k) = R_{12}{}'(w_k)\left[P_{11}(w_k)\right]^{-1}. \tag{14}$$

The following are the multivariate versions of Theorem 1.

Theorem 2. The estimator in (14) is a consistent estimator of $Q_{21}Q_{11}^{-1}$ for each fixed k.

As before, a fixed number of frequencies can be combined in a multivariate regression to give a consistent estimate.

Theorem 3. Using the transformations in (13), the ordinary least squares estimator $\hat{\beta}_n$ of $\underline{\beta}$ is a consistent estimator for $Q_{21}Q_{11}^{-1}$ using the regression model

$$Y_k = \beta X_k + E_k,\; k = 1,2,\ldots,m$$

where Y_k and X_k are as before and the estimator is defined as

$$\hat{\beta}_n = \left(\sum_{k=1}^{m} Y_k{}' X_k\right)\left(\sum_{k=1}^{m} X_k{}' X_k\right)^{-1}.$$

5. AN EXAMPLE AND SIMULATIONS

A. Demonstration with Real Data: As an example, we use quarterly US consumption and income data set of Beaulieu and Miron (1993). It covers the period 1946:1 through 1985:4. Data are log transformed and seasonally adjusted using X12 adjustment program. First of all, we check if the series are of $I(1)$. The time series plots and their identification plots are given in Figure 1 below.

Both series are modelled with first order autoregressive model, AR(1) as suggested by the values of AIC and SBC statistics obtained from PROC ARIMA in SAS. The autocorrelations decay very slowly. Here, X and Y denote income and consumption under the logarithmic transformation, respectively. First differenced series and their identification plots are given in Figure 2.

The stationarity of the series have been checked with standard Dickey-Fuller test and the periodogram based unit root test proposed by Akdi and Dickey (1998). The models for X and Y are

$$X_t = \alpha_{X,0} + \alpha_{X,1} X_{t-1} + e_{X,t} \text{ , } t = 1,2,3,...,160$$

$$Y_t = \alpha_{Y,0} + \alpha_{Y,1} Y_{t-1} + e_{Y,t} \text{ , } t = 1,2,3,...,160$$

and the results heve been summarized below:

Table 1. Summary of the Results

	$\hat{\tau}_\mu$	Critical Value	Periodogram	$T_n(w_1)$	Critical Value	Conclusion
X_t	3.201	-2.89	66.9539	688.254	0.178	Unit Root
Y_t	2.415	-2.89	50.6841	300.582	0.178	Unit Root
∇X_t	-13.442	-2.89	0.0040204	0.03953	0.178	Stationary
∇Y_t	-14.811	-2.89	0.0041481	0.024176	0.178	Stationary

The periodogram test is left tailed. We fail to reject the null hypothesis of a unit root for each series. That is, both series are integrated of order 1 according to the results of Dickey-Fuller and periodogram tests. The periodiogram based method has certain advantages over conventional tests. Firstly, conventional tests require the estimation of too many AR parameters to account for the dynamics/seasonality of the series. Secondly, test results change with the sample size in conventional tests, while the periodogram based method requires no parameter estimation except for variance. Thirdly, the critical values of the test statistics are free of sample size constraints. Thus, these might have considerable advantages, especially for small samples.

In order to check whether these series are cointegrated, we calculate the periodograms and the real part of the cross periodograms of the series and regress the real part of the cross periodograms on the periodogram of one of the series (income). The estimate of the regressioon coefficient with spectral regression is $0.86814 \approx 0.87$. The time series and its identification plots of $Z_t = Y_t - 0.87 X_t$ given in Figure 3, seem to show a stationary series but this requires a statistical check. If this series is stationary, then the estimated cointegration vector is $\underline{\beta} = (-0.87, 1)'$. The values of AIC statistic and the identification plots imply an AR(1) model for Z_t, $Z_t = \alpha Z_{t-1} + \upsilon_t$. To test $H_0 : \alpha = 1$ against $H_a : |\alpha| < 1$ one can regress $D_t = Z_t - Z_{t-1}$ on Z_{t-1} and calculate the standard t-statistic as: $\hat{\tau}_a = \hat{\alpha} / s(\hat{\alpha}) \approx -3.325$ which is smaller than 10% critical value -3.134 (obtained by simulation- see Table 2) and we reject H_0 at the 10% level. The parameter estimates and their standard errors are given below:

$$
\begin{array}{lccc}
\nabla Z_t & = \hat{\alpha}_0^* & + \quad \hat{\alpha}_1^* & Z_{t-1} \\
Est. & -0.0192 & -0.104 & \\
Std.Err. & 0.0056 & 0.0313 & \\
t-stat & -3.428 & -3.325 &
\end{array}
$$

The power is approximated as 0.804 which is obtained in a similar way given in Table 3 in the Annex.

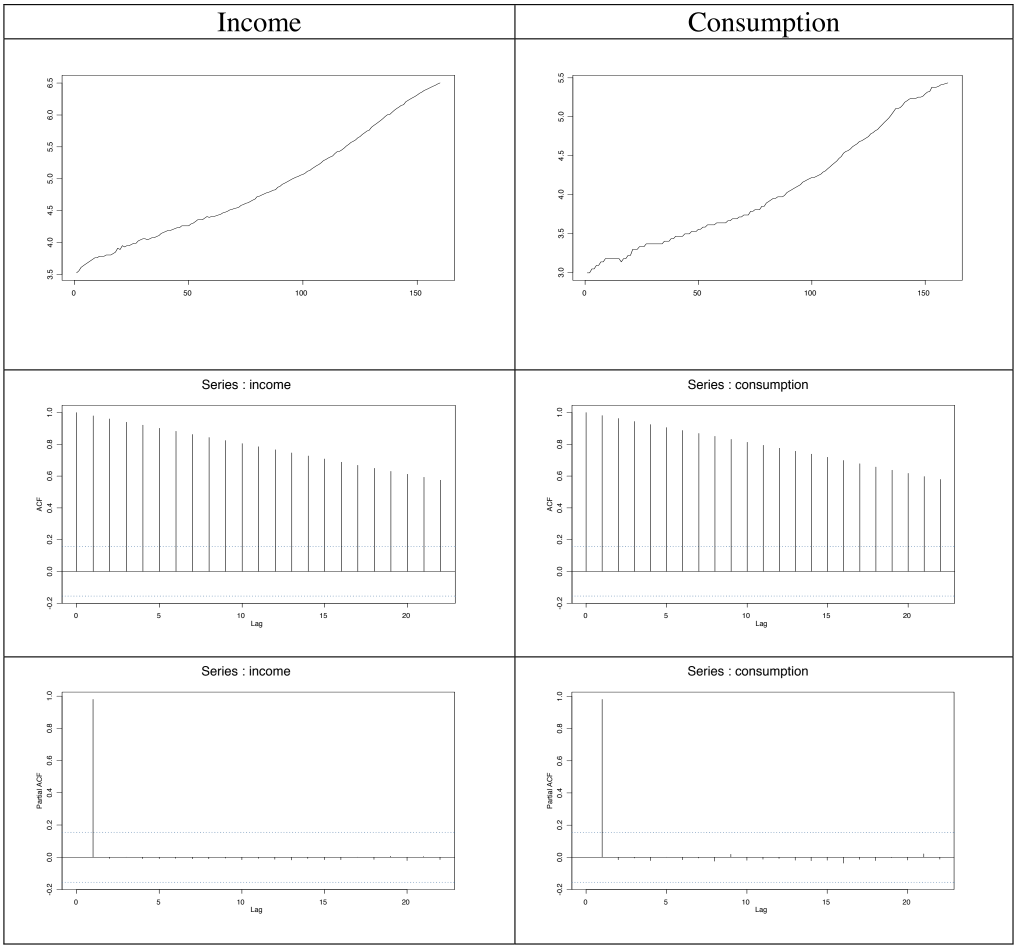

Figure 1. Original Series and Their Identification Plots

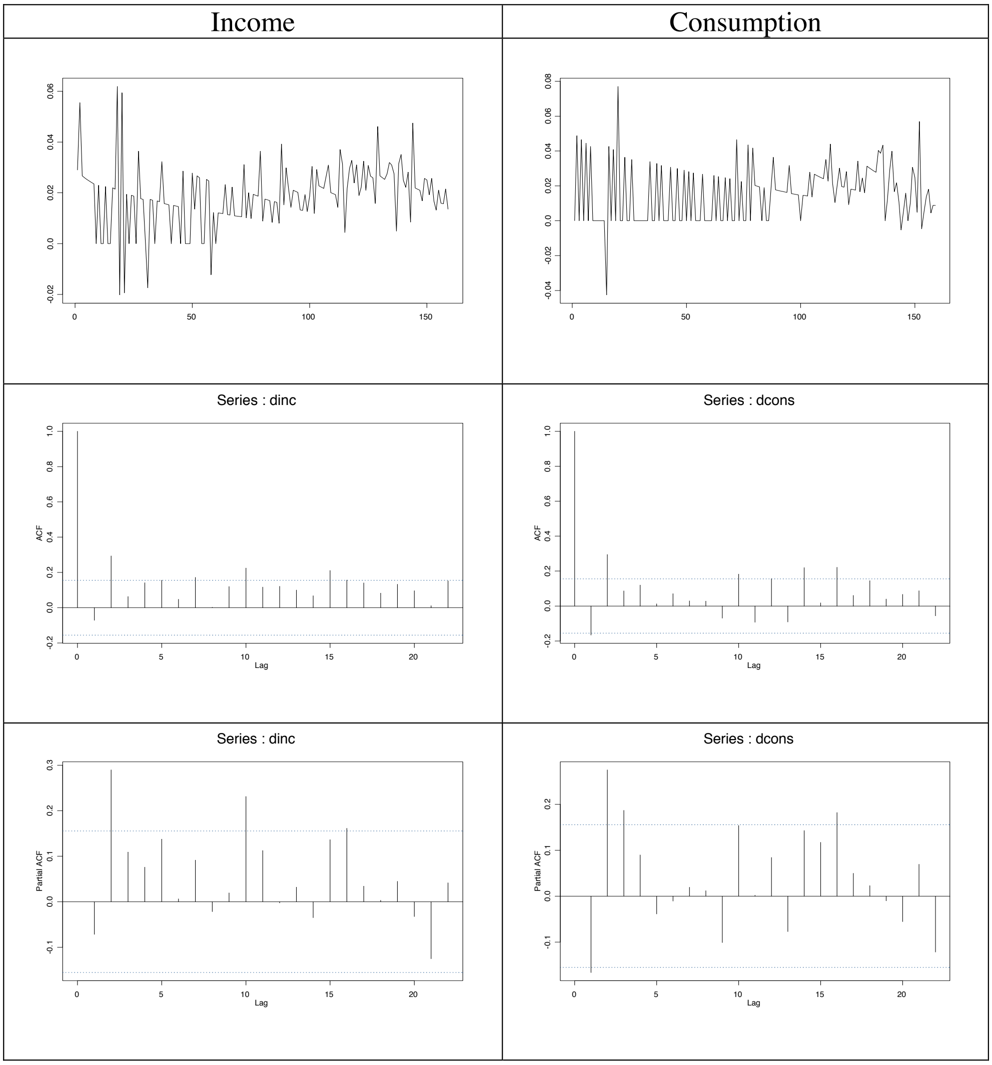

Figure 2. First Differences and Their Identification Plots

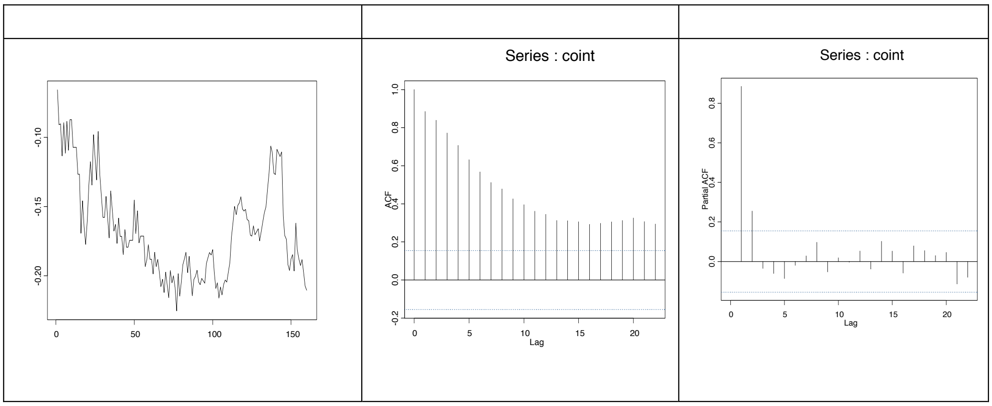

Figure 3. Cointegrated Series

This critical value comes from a simulation reported in Table 2. Bivariate series with roots 1 and ρ were generated, Z was computed by the periodogram method shown here and then the differenced Z was regressed on its first lag Z and empirical percentiles computed. The percentiles of τ_a can be used as the critical values for testing the null hypothesis of no cointegration. The table is similar to the table of Engle and Granger but here the periodogram method is used rather than ordinary regression to construct Z. In other words, adjusted consumption and income are cointegrated at 10% significance level. In a similar way, the performance of the test is tabulated for different ρ's as listed in Table 3. Here, 1000 replications are used and for different ρ's the number of rejections were counted.

In addition, the cointegrating relationship between consumption and income is analyzed with the Engle and Granger approach. The estimated cointegrating vector $(0.856996 \approx 0.86)$ is very close to the one estimated through the spectral regression approach. When the first differenced residual series R_t (obtained from the regression of Y_t on X_t) is regressed on R_{t-1}, we get the following estimation results:

$$
\begin{array}{llll}
\nabla \hat{R}_t & = & -0.000684 & -\ 0.10886\,\hat{R}_{t-1} \\
Std.Err & & (0.0011) & (0.0327) \\
t-stat & & -0.620 & -3.326
\end{array}
$$

Since, $\hat{\tau}_\mu(E-G) = -3.326 < -3.033$, we reject the null hypothesis of no cointegration at the 10% level. The critical value -3.033 is obtained by simulation with 5000 replications. Note that both tests fail to reject the null hypothesis of no cointegration at 5% significance level.

Moreover, the Johansen (1988) method is also applied for investigating the cointegration relation between these variables. The corresponding squared canonical correlations are 0.099310 and 0.050393. Then, the value of Johansen's trace statistics is calculated as $\hat{\lambda}_{trace} = -160[\ln(1-0.099310) + \ln(1-0.050393)] = 25.008$ which is greater than 10% critical value (10.3) and we reject the null hypothesis of no cointegration. Thus, we find that the consumption and income series are cointegrated at 10% significance level.

B. Simulation Study: We consider the series generated from a normal distribution with mean zero and variance 1 according to (15). That is, $e_{1,t} \sim N(0,1)$, $e_{2,t} \sim N(0,1)$ with $U_t = \rho_1 U_{t-1} + e_{1,t}$ and $S_t = \rho_2 S_{t-1} + e_{2,t}$ for $\rho_1 = 1$, $|\rho_2| < 1$. The equations in (2) allow us to write

$$
\begin{aligned}
Z_{1,t} &= U_t + 2S_t \\
Z_{2,t} &= 3U_t + 4S_t
\end{aligned}
\qquad (15)
$$

where U_t represent a unit root time series and S_t represents a stationary time series. Notice that both series $Z_{1,t}$ and $Z_{2,t}$ are nonstationary because both include U_t. That is, the nonstationary bivariate series can be written as

$$\begin{bmatrix} Z_{1,t} \\ \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ \\ 3 & 4 \end{bmatrix} \begin{bmatrix} U_t \\ \\ S_t \end{bmatrix}. \tag{16}$$

But $Z_{2,t} - 3Z_{1,t} = -2S_t$ which is stationary. Based on this cointegration relationship, in the following simulations we expect to get estimates close to 3 for different numbers of observations. We run 10,000 regressions for each case (different ρ_2 and n) and average these spectral regression estimates of the cointegrating vectors. We also run 10,000 regressions of $Z_{2,t}$ on $Z_{1,t}$ for each case and average these 10,000 ordinary least squares estimates of the cointegrating coefficient as in Engle and Granger (1987), labeled OLS in Table 1a. All frequency regressions does not have fixed k as n gets large so our asymptotic theory does not apply to it. Different number of frequencies were considered in the regressions and we observed that using all frequencies in the regression gives the worst result. We also analyze the bivariate series with Johansen's method. Standard deviations and mean squared errors of estimation are reported for comparison. For the bivariate case, the sample sizes used as $n = 50, 100, 200$. The values of the parameters $\rho_1 = 1$ and $\rho_2 = 0.7, 0.8, 0.9, 0.95, 0.99$ are considered and the results are tabulated in Table 1b. Notice that there is no cointegration when $\rho_2 = 1$. In our simulations, we observe that Johansen's method produces some bizarre outliers which cause large mean squared error and standard deviations and therefore the bias of Johansen's method is large.

In Table 1a, we take the average of 10,000 regression estimates obtained from both ordinary least squares and spectral regression for different values of ρ_2. We are looking for the values close to 3. Using all frequencies in the regression yields the worse results. Even in the worst case spectral regression gives a better result than that obtained from ordinary least squares. As it is seen from the table, the estimates gets further away from 3 as ρ_2 approaches 1. In Table 1b, SRE is our periodogram estimator using all frequencies in the regression. When ρ_2 takes the value 1, there is no cointegration because the matrix Π has rank of zero.

6. CONCLUSION

In this study, periodogram based cointegration method have been proposed. The method improves the OLS method proposed by Engle and Granger (1987) in certain cases.

7. REFERENCES

Akdi, Y., Dickey, D. A., 1998. Periodograms of Unit Root Time Series: Distributions and Tests, Communications in Statistics: Theory and Methods, 27, 69-87.

Beaulieu, J., Miron, J. A., 1993. Seasonal Unit Roots in Aggregate US Data, Journal of Econometrics, 55, 305-328.

Bloomfield, P., 1976. Fourier Analysis of Time Series: An Introduction, Wiley, New York.

Boswijk, H. P., Lucas, A., 2002. Semi-nonparametric Cointegration Testing, Journal of Econometrics, 108, 253-280.

Breitung, J., 2002. Nonparametric Tests for Unit Roots and Cointegration, Journal of Econometrics, 108, 343-368.

Chambers, M. J., 2001. Temporal Aggregation and the Finite Sample Performance of Spectral Regression Estimators in the Cointegrated System: A Simulation Study, Econometric Theory, Vol. 17, Number 3, 591-607.

Chen, W. W., Hurvich, C. M., 2003. Estimating Fractional Cointegration in the Presence of Polynomial Trends, Journal of Econometrics, 117, 95-121.

Choi, I., Phillips, P. B. C., 1993. Testing for a Unit Root by Frequency Domain Regression, Journal of Economics, 59, 263-286.

Deo, R. S., Hurvich, C. M., 2001. On the log Periodogram Regression Estimators of the memory Parameters in the Long Memory Volatility Models, Econometric Theory, Vol. 17, Number 4, 686-710.

Dickey, D. A., Fuller, W. A., 1979. Distributions of Estimators for Autoregressive Times Series with a Unit Root, Journal of American Statistical Association, 74, 427-431.

Dickey, D. A., Pantula, S. G., 1987. Determining the order of Differencing in Autoregressive Processes, Journal of Business and Economics Statistics, 5, 455-461.

Dickey D. A., Hasza, D. P., Fuller, W. A., 1984. Testing For Unit Roots in Seasonal Time Series, Journal of American Statistical Association, 79, 355-367.

Engle, R.F., C.W.J. Granger., 1987. Cointegration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 251-276.

Fuller, W.A., 1996. Introduction to Statistical Time Series, Wiley, New York.

Hylleberg, S., Engle, R. F., Granger, C. W. J., Yoo, B. S., 1990. Seasonal Integration and Cointegration, Journal of Econometrics, 99, 215-238.

Johansen, S., 1988. Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamics and Control, 12, 231-254.

Levy, D., 2002. Cointegration in Frequency Domain, Journal of Time Series Analysis, Vol.23, No.3, 333-339.

Marunicci, D., 2000. Spectral regression For Cointegrated Time Series With Long-Memory Innovations, Journal of Time Series Analysis, Vol. 21, Issue 6, 685-705.

Stock, J. H., 1987. Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors, Econometrica, 55, 1035-1056.

Stock, J. H., Watson, M. W., 1993. A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, Econometrica, 61, 783-820.

EŞBÜTÜNLEŞME İÇİN TAHMİN VE TEST: SPEKTRAL BİR REGRESYON YAKLAŞIMI

ÖZET

Ekonometri ve zaman serileri alanındaki popüler bir konu, vektör otoregressif zaman serilerinin bileşenleri arasındaki eşbütünleşme ilişkisidir. Engle ve Granger (1987)'in çalışmalarından sonra problem önemli hale gelmiş ve Johansen (1988), Stock ve Watson gibi başka pek çok yazar tarafından da bu probleme işaret edilmiştir. Engle ve Granger'in en küçük kareler metodu ile Johansen'in koşullu maksimum olabilirlik metodu en çok dikkati çekenlerdendir. Bu testler ekonomik zaman serilerine rutin olarak uygulanmıştır çünkü eşbütünleşme nosyonu doğal bir yoruma sahiptir. Bizim metodumuz, eşbütünleşmiş zaman serileri arasındaki eşbütünleşme ilişkisini tahmin etmek için çapraz periyodogramın düşük frekanslı bileşenlerini kullanır. Bazı durumlarda bu metod, Engle ve Granger tarafından önerilen sıradan en küçük kareler metodunun sonuçlarını geliştirir.

Anahtar Kelimeler: Zaman serileri, Eşbütünleşme, Periyodogram ordinat, Spektral regresyon.

Annex. Tables

Table 1a. Simulations (* marks the closest average to 3 in each row)

	5 Freq	10 Freq.	All Freq.	OLS
n=50				
$\rho_2 = 0.7$	2.641*	2.619	2.600	2.482
$\rho_2 = 0.8$	2.524*	2.511	2.498	2.421
$\rho_2 = 0.9$	2.273	2.367*	2.362	2.329
$\rho_2 = 0.95$	2.282*	2.280	2.279	2.268
$\rho_2 = 0.99$	2.221*	2.221	2.220	2.217
n=100				
$\rho_2 = 0.7$	2.843*	2.822	2.786	2.617
$\rho_2 = 0.8$	2.736*	2.715	2.684	2.552
$\rho_2 = 0.9$	2.536*	2.523	2.507	2.436
$\rho_2 = 0.95$	2.380*	2.373	2.367	2.337
$\rho_2 = 0.99$	2.235*	2.234	2.233	2.231
n=200				
$\rho_2 = 0.7$	2.957*	2.948	2.916	2.746
$\rho_2 = 0.8$	2.905*	2.890	2.852	2.687
$\rho_2 = 0.9$	2.746*	2.727	2.693	2.569
$\rho_2 = 0.95$	2.546*	2.533	2.514	2.446
$\rho_2 = 0.99$	2.267*	2.267	2.265	2.260

Table 1b. Simulations (Comparisons for SRE with OLS and Johansen with respect to MSE and STD, used all frequencies in SRE)

N=50	$\rho_2 = 0.7$	$\rho_2 = 0.8$	$\rho_2 = 0.9$	$\rho_2 = 0.95$	$\rho_2 = 0.99$
SRE	2.600	2.498	2.362	2.279	2.220
MSE(SRE)	0.245	0.344	0.510	0.626	0.715
STD(SRE)	0.292	0.305	0.321	0.325	0.327
OLS(EG)	2.482	2.421	2.329	2.268	2.217
MSE(EG)	0.311	0.381	0.310	0.238	0.247
STD(EG)	0.205	0.215	0.310	0.238	0.247
JOH	3.632	2.408	3.286	8.48	2.449
MSE(JOH)	3482.5	7081.5	833.22	36382835	2427.5
STD(JOH)	59.01	84.15	28.87	603.187	49.268
N=100					
SRE	2.786	2.684	2.507	2.367	2.233
MSE(SRE)	0.097	0.169	0.333	0.500	0.693
STD(SRE)	0.227	0.264	0.300	0.315	0.326
OLS(EG)	2.617	2.552	2.436	2.337	2.231
MSE(EG)	0.178	0.238	0.364	0.491	0.652
STD(EG)	0.179	0.194	0.215	0.228	0.246
JOH	2.968	2.936	-70.28	1.0248	2.178
MSE(JOH)	193.68	257.95	52078513	17247	1610.06
STD(JOH)	13.91	16.06	7216.53	131.319	40.119
N=200					
SRE	2.916	2.852	2.693	2.514	2.265
MSE(SRE)	0.026	0.057	0.161	0.324	0.649
STD(SRE)	0.168	0.188	0.258	0.297	0.332
OLS(EG)	2.746	2.687	2.569	2.446	2.260
MSE(EG)	0.084	0.124	0.223	0.353	0.608
STD(EG)	0.141	0.162	0.193	0.216	0.248
JOH	3.014	3.051	3.527	9.018	2.789
MSE(JOH)	0.021	8.733	3028.01	35054277	14941.76
STD(JOH)	0.144	2.95	55.03	592.06	122.24

Table 2. Critical values for τ_a (5,000 replicates) The values for $\rho = 1$ are used for testing the null hypothesis of no cointegration.

	0.01	0.05	0.10	0.90	0.95	0.99
N=50						
$\rho = 1$	-4.121	-3.464	-3.173	-0.885	-0.488	0.290
$\rho = 0.95$	-4.225	-3.613	-3.264	-1.203	-0.880	-0.300
$\rho = 0.90$	-4.318	-3.685	-3.336	-1.412	-1.148	-0.593
$\rho = 0.80$	-4.498	-3.906	-3.610	-1.774	-1.530	-1.079
$\rho = 0.70$	-4.788	-4.164	-3.873	-2.065	-1.851	-1.419
N=100						
$\rho = 1$	-4.125	-3.434	-3.114	-0.967	-0.602	0.188
$\rho = 0.95$	-4.183	-3.659	-3.370	-1.523	-1.255	-0.677
$\rho = 0.90$	-4.420	-3.938	-3.655	-1.936	-1.714	-1.226
$\rho = 0.80$	-5.146	-4.513	-4.233	-2.563	-2.368	-1.982
$\rho = 0.70$	-5.650	-5.032	-4.737	-3.031	-2.835	-2.489
N=160						
$\rho = 1$	-4.050	-3.415	-3.134	-1.005	-0.588	0.159
$\rho = 0.95$	-4.330	-3.810	-3.544	-1.820	-1.605	-1.133
$\rho = 0.90$	-4.815	-4.305	-4.043	-2.426	-2.231	-1.823
$\rho = 0.80$	-5.726	-5.183	-4.924	-3.318	-3.132	-2.735
$\rho = 0.70$	-6.432	-5.880	-5.629	-3.969	-3.796	-3.403
N=200						
$\rho = 1$	-3.994	-4.424	-3.121	-1.047	-0.646	0.105
$\rho = 0.95$	-4.446	-3.925	-6.643	-1.995	-1.781	-1.317
$\rho = 0.90$	-5.022	-4.544	-4.278	-2.703	-2.502	-2.107
$\rho = 0.80$	-6.043	-5.538	-5.307	-3.734	-3.526	-3.099
$\rho = 0.70$	-6.807	-6.361	-6.103	-4.494	-4.261	-3.848

Table 3. Power of τ_a (number of rejections out of 1,000 replications)

$\alpha = 0.05$				$\alpha = 0.10$			
$\rho \downarrow N \rightarrow$	100	160	200	$\rho \downarrow N \rightarrow$	100	160	200
1.00	0.041	0.045	0.047	1.00	0.107	0.093	0.103
0.95	0.087	0.133	0.169	0.95	0.160	0.213	0.290
0.90	0.167	0.329	0.516	0.90	0.287	0.533	0.705
0.80	0.452	0.866	0.960	0.80	0.649	0.950	0.978
0.70	0.734	0.984	1.000	0.70	0.868	0.996	1.000
0.60	0.886	1.000	1.000	0.60	0.964	1.000	1.000
0.50	0.965	1.000	1.000	0.50	0.992	1.000	1.000