Periodic Properties of Interpolated Time Series*

by

Hashem Dezhbakhsh and Daniel Levy
Department of Economics
Emory University
Atlanta, GA 30322-2240
Tel: (404) 727-2941
Fax: (404) 727-4639

October 25, 1993

Abstract:

Although linearly interpolated series are often used in economics, little has been done to examine the effects of interpolation on time series properties and on statistical inference. We show that linear interpolation of a trend stationary series superimposes a 'periodic' structure on the moments of the series. Using conventional time series methods to make inference about the interpolated series may therefore be invalid. Also, the interpolated series may exhibit more shock persistence than the original trend stationary series.

---

1. Introduction

Analysis of economic time series is often limited by the sampling frequency. For example, aggregate capital stock data is usually collected annually, while economists often require quarterly data. Also, some prewar U.S. output and price series are only available for benchmark years which may be a decade apart, but economists who study prewar business cycles need at least annual data. A technique often used to increase data frequency or to estimate missing observations is segmented linear interpolation.[1] It involves generating new data points along each segmented line connecting two available and consecutive data points. Generated data are sometimes padded by adding a serially correlated term derived from another indicator or component series. Examples include widely used series constructed by Kuznets (1946, 1961), Shaw (1947), Friedman and Schwartz (1982), and Romer (1989).

As Romer (1989) argues, the accuracy of our views about prewar cycles depends on the quality of the interpolated series because many studies of important macroeconomic relationships use these series. Yet the effect of interpolation on the stochastic properties of time series has hardly been examined. Furthermore, little is known about the way interpolation affects the statistical measures computed from the series in question. For example, Stock and Watson (1986) conjecture that linear interpolation may have caused the difference between the GNP shock persistence of the prewar and postwar series, as reported in their study and in studies by Campbell and Mankiw (1987) and DeLong and Summers (1988).

Jaeger's (1990) Monte Carlo results suggest that segmented linear interpolation reduces the size of shock persistence in a difference stationary series. In this paper we analytically examine the effect of segmented linear interpolation on the time series properties of a trend stationary process.[2] We compare moments and cross moments of the original and interpolated series and discuss consequences of applying conventional time series methods to interpolated series. In particular, we discuss implications for testing random walk vs trend stationarity using shock

---

[1] It is interesting to note that linear interpolation was apparently used as early as 2000 B.C. in ancient Babylonian Mesopotamia to construct exponential tables. See Boyer and Merzbach (1991).

[2] By considering a trend stationary series we extend Jaeger's (1990) work. In addition, the findings reported in Cochrane (1988), Christiano and Eichenbaum (1990), and Rudebusch (1993) that indicate the unit root component of GNP data may be smaller than previously thought provide an added incentive for considering a trend stationary series.

persistence measures.

The paper is organized as follows. In the next section we model a linearly interpolated series and evaluate its properties. In section 3 an illustrative example is presented to highlight the effect of interpolation on moments of the process; practical implications are also discussed. Section 4 concludes the paper.

2. Linear Interpolation of Trend Stationary Time Series

We start with a trend stationary series and rewrite it by introducing a new notation. This notation is useful for analyzing periodic and interpolated series, and thus is more suitable for our purpose. We then divide the series into segments of equal length, drop all but one observation in each segment, and reconstruct the missing observations by linear interpolation. The original and interpolated series are compared and their properties are examined.

Consider the following trend stationary model characterizing an economic time series:

$$Y_T - \ T = U_T, \qquad\qquad\qquad T = 1, 2, \ldots, \tag{1}$$

where $\ T$ is the deterministic component of the series, and U_T is the stochastic component with a stationary autoregressive-moving average (henceforth ARMA) structure. We assume U_T is ARMA(1, 1),

$$U_T = \frac{1 + \ L}{1 - \ L}\ _T, \tag{2}$$

where $\ _T$ is i.i.d. $(0, \ ^2)$, $|\ |$ and $|\ | < 1$, and L is the standard lag operator with $L\ _T = E_{T-1}$. An ARMA(1, 1) process is adopted for the sake of concreteness. Using an ARMA(p, q) process introduces more complexities but does not change our findings (see also footnote 6).

The above notation is adequate for modeling annual series where T denotes the year. However, when applied to a 'periodic' (e. g., quarterly or monthly) series, the single index T that identifies the period becomes merely a counting device. The setting considered here requires a more specific notation. Therefore, we write the Y_T series as $\{y_{t,\ }\}$, with $i = 1, 2, \ldots, s$, and

$t = 0, 1, 2, \ldots$, where each element of the $y_{t,\,i}$ series corresponds to an element of Y_T as follows: $y_{0,1} = Y_1$, $y_{0,2} = Y_2$, $\ldots$, $y_{0,s-1} = Y_{s-1}$, $y_{0,s} = Y_s$, $y_{1,1} = Y_{s+1}$, $\ldots$, and $y_{t,\,i} = Y_{st+i}$. Here, 'period' refers to the time span covered by s subperiods at any t.[3]

We now introduce a new lag operator L, which we call *periodic lag operator*. The operator follows the rule

$$
Ly_{t,\,i} \quad \begin{cases} y_{t,\,i-1} & \text{if } i > 1 \\ y_{t-1,\,s} & \text{if } i = 1.[4] \end{cases}
$$

Then, an alternative but observationally equivalent representation of (1) and (2) is given by

$$
y_{t,\,i} - \quad (st + i) = \frac{1 + L}{1 - L} \quad _{t,\,i}, \tag{3}
$$

where $_{t,\,i}$ and $_T$ have the same correspondence as $y_{t,\,i}$ and Y_T. The deterministic component of the process is now $(st + i)$.

The detrended $y_{t,\,i}$ series only contains the stochastic component which we denote by $\bar{y}_{t,\,i}$. The first two moments, cross moments, and the autocorrelation function of $\bar{y}_{t,\,i}$ are given by

$$
\mu_{\dot{y}} = E(\bar{y}_{t,\,i}) = 0, \tag{4a}
$$

$$
_{y}(0) = Var(\bar{y}_{t,\,i}) = \quad ^2\left(\frac{1 + \quad^2 + 2}{1 - \quad^2}\right), \tag{4b}
$$

$$
_{y}(1) = Cov(\bar{y}_{t,\,i}, L\bar{y}_{t,\,i}) = \quad ^2\left[\frac{(1 + \quad)(\quad + \quad)}{1 - \quad^2}\right]^{1}, \tag{4c}
$$

$$
_{y}(k) = \quad _{y}(k-1) \quad \text{for } k \quad 2, \text{ and} \tag{4d}
$$

---

[3]Note that this notation is general enough to fit any periodic structure or data frequency. For example, t and i may, respectively, denote decades and years within a decade. In fact, t may denote any arbitrary time interval, in which case i denotes nonoverlapping and equal subperiods within that interval. In a more conventional setting for quarterly data, t and i denote years and quarters, respectively.

[4]Obviously, $L^s y_{t,\,i}$ equals $y_{t-1,\,i}$. Also, note that applying the periodic lag operator to the indices themselves follows the same rule. That is, for $i > 1$, $Li = i - 1$ and $Lt = t$, while for $i = 1$, $Li = s$ and $Lt = t - 1$.

$$r_y(k) = \frac{\gamma_y(k)}{\gamma_y(0)}.$$

(4e)

The process is stationary and its moments do not depend on i.

Now consider a situation where in each period only one of the s data points is available; this point pertains to the end of the period, $y_{t, s}$. To increase frequency, missing observations are generated by segmented linear interpolation. Each segment covers one period. The $s-1$ missing observations in that period are generated according to the following linear interpolation:

$$x_{t, i} \quad \frac{i}{s}\left(y_{t, s} - y_{t-1, s}\right) + y_{t-1, s} = \frac{i}{s}\, y_{t, s} + \frac{s-i}{s}\, y_{t-1, s}.$$

(5)

We refer to $y_{t, i}$ as the original series and to $x_{t, i}$ as the interpolated series. Obviously, $x_{t, i} = y_{t, i}$ for $i = s$, as the interpolation does not change the available observations.[5]

Substituting $y_{t, s}$ and $y_{t-1, s}$ from (3) into (5) yields

$$\tilde{x}_{t, i} = x_{t, i} - \quad (st + i) = \frac{i}{s}\left(\frac{1+}{1-}\frac{L}{L}\right)_{t, s} + \frac{s-i}{s}\left(\frac{1+}{1-}\frac{L}{L}\right)_{t-1, s}.$$

(6)

It can be seen from (6) that the interpolated series has the same deterministic component as the original series; this is obviously an appealing feature of linear interpolation. The mean of the stochastic component of the interpolated series also remains zero.

However, the interpolation does alter the variance/covariance structure of the series. Using (5) and (6) it is straightforward to show that the variance and autocovariance functions of the interpolated series are respectively

$$\gamma_{\tilde{x}, i}(0) = \frac{i^2 + (s-i)^{2^1}}{s^2}\, \gamma_y(0) + \frac{2i(s-i)}{s^2}\ ^3\ \gamma_y(1), \qquad \text{and}$$

(7)

$$\dot{x}_{,i}(k) = \begin{cases} \dfrac{s(s-i)+(2i-s)(i-k)}{s^2} \; _y(0) + \dfrac{si-(2i-s)(i-k)}{s^2} \; _3 \; _y(1), & \text{if } 0 < k \quad i \\[2em] \left[\dfrac{s(s-i)+(2i-s)(L^k i)}{s^2} \; _{-1} + \dfrac{si-i(L^k i)}{s^2} \; _3 \right] \; _4\left\lceil\frac{k-i}{s}\right\rceil \; _y(1) - \dfrac{(i-s)(L^k i)}{s^2} A, & \text{if } k > i \end{cases}$$

(8)

where $_y(0)$ and $_y(1)$ are given in (4), $\left\lceil\dfrac{k-i}{s}\right\rceil$ is the closest integer larger than or equal to $\dfrac{k-i}{s}$, and $A = \; _y(0)$ if $\left\lceil\dfrac{k-i}{s}\right\rceil = 1$, or else $A = \left(\; _4\left\lceil\frac{k-i}{s}\right\rceil -5 \right) \; _y(1)$.[6] In $L^k i$, the periodic operator L is applied k times to the index i. The autocorrelation function of the interpolated process is given by

$$r_{\dot{x},i}(k) = \frac{\dot{x}_{,i}(k)}{\sqrt{\dot{x}_{,i}(0) \; \dot{x}_{,L^k i}(0)}}.$$

(9)

It is apparent from equations (7)–(9) that the variance, covariance, and autocorrelation functions of the interpolated series are invariant to changes in t but they vary with i—such variations are termed periodic. These functions also differ from the variance and autocovariance functions of the original series given in (4). The magnitude of such periodic variations and differentials depends on parameters of the stochastic component of the original series.

3. An Example and Interpretation of Results

The following example illustrates the periodic variation in the theoretical variance and autocorrelations of the interpolated series. It also shows the extent of divergence between these moments and the corresponding moments of the original series. The moments are computed using equations (4), (7), and (8) where $_y(0)$ and $_y(1)$ are obtained assuming the stochastic component of the original series to be AR(1) with parameters ^{2} and $.$[7] We set $^2 = 1$ and $s = 4$. The results are displayed in Figures 1–3.

Figure 1 shows the variance of the interpolated series for various values of $.$ Three distinct

---

[6]Similar expressions can be derived for $k < 0$. Details of the derivations are available upon request. It is also noted that similar expressions can be derived for an ARMA(p, q) process. For example, the variance in case of a general ARMA(p, q) process would be $_{\dot{x},i}(0) = \dfrac{i^2 +(s-i)^{2^1}}{s^2} \; _y(0)+ \dfrac{2i\,(s-i)}{s^2} \; _y(4).$

[7]We choose AR(1) for expositional convenience and also because of its popularity in the 'unit root vs trend stationary' literature. Our findings, however, are robust to alternative specifications.

values are observed, one for $i = 4$, one for $i = 2$, and one for $i = 1$ or 3. As mentioned earlier, $\ddot{x}_{t,4} = \ddot{y}_{t,4}$ and therefore $\gamma_{\ddot{x},4}(0)$ equals the variance of the original series (solid line). This variance is larger than the other two periodic variances by as much as 100%. Figure 2 shows the first order autocorrelation of the interpolated and the original series (solid line) for various values of ϕ.[8] Here we have two distinct periodic autocorrelations, and both are considerably larger than the autocorrelation of the original series regardless of the value of ϕ. To examine the behavior of the periodic autocorrelations of the interpolated series at higher lags, we set $\phi = 0.9$ and compute $r_y(k)$ and $r_{\ddot{x},i}(k)$, $i = 1$ to 4, and $k = 1, 2, \ldots, 8$.[9] The results, displayed in Figure 3, indicate that there are two or three distinct periodic autocorrelations at each k. These parameters are larger than the autocorrelation of the original series (solid line).

The above example illustrates that the moments of a linearly interpolated series exhibit significant periodic variations even though the original series is stationary around its trend. The number of distinct periodic parameters depends on the interpolation frequency. For example, if the interpolation is intended to quadruple the frequency, the number of periodic moments will be 2–3 times the number of original moments. Moreover, the periodic autocorrelations of the interpolated series are larger than the autocorrelations of the original series at matching lags.

The theoretical results presented here have important practical implications. First, segmented linear interpolation alters the stochastic properties of a trend stationary series by introducing periodicity into moments of its stochastic components. This indicates nonstationarity if viewed in a conventional sense. However, the interpolated series is *periodically* stationary because moments are invariant to changes in the period despite their variation within the period.[10] Ignoring this periodicity and using conventional time series analysis may lead to dynamic misspecification, and consequently to invalid statistical inference. For example, standard unit root tests may not be valid

---

[8] In the present example the original detrended series is AR(1), so its k-th order autocorrelation is ϕ^k, $k \geq 0$.

[9] This particular value of ϕ is chosen because the stochastic component of economic series is expected to have a high serial correlation. For the sake of completeness, we repeated this exercise for smaller values of ϕ and found the variations reported here to be even more pronounced.

[10] The moment parameters of a periodically stationary series do not change across periods but they change within a period. For example, in a quarterly setting there are four sets of moments pertaining to each quarter and these moments are invariant to a change in the year. Periodic processes have recently received considerable attention. See, e. g., Osborn (1988, 1991), Barsky and Miron (1989), and Chatterjee and Ravikumar (1992).

when applied to linearly interpolated series. Instead, such series should be analyzed using the tools developed for periodically stationary series. See, e. g., Osborn (1988, 1991) and the references therein.

Second, the autocorrelation function of a trend stationary series appears to lie below the periodic autocorrelation functions of its linear interpolation. Therefore, standard estimates of the original autocorrelation function using the interpolated series will be biased upward. Consequently, any shock persistence measure that is a linear combination of the estimated autocorrelations, as in Cochrane (1988), may be positively biased, leading to an overestimation of the random walk component. Any evidence of a random walk found in interpolated series should therefore be interpreted with caution.

4. Conclusions

Our results suggest that segmented linear interpolation of a stationary process leads to varying moments that may be viewed as an indication of nonstationarity in a conventional sense. The variation in moments is periodic, however, and the interpolated series are in fact periodically stationary. Therefore, such series should be analyzed in the context of periodic time series models rather than by conventional methods.

Furthermore, ignoring the periodic nature of the interpolated series and estimating the autocorrelations by conventional methods results in estimates which are biased upward. Consequently, shock persistence measures computed from such estimates may also be biased upward, leading to an incorrect inference that the series contains a random walk.

Jaeger's (1990) Monte Carlo results suggest that interpolation introduces a downward bias in the shock persistence measures if the original series follows a random walk. We, on the other hand, find an upward bias if the original series is trend stationary. This implies that the direction of the bias caused by linear interpolation will depend on whether the original series follows a random walk or a trend stationary process. This indeterminacy in the direction of the bias makes any random walk vs trend stationarity inference about a linearly interpolated series suspect.

References

Barsky, R.B. and J.A. Miron, 1989, The seasonal cycle and business cycle, Journal of Political Economy 97, 503–534.

Boyer, C.B. and U.C. Merzbach, 1991, A history of mathematics, 2nd ed. (Wiley, New York).

Campbell, J.Y. and N. G. Mankiw, 1987, Are output fluctuations transitory?, Quarterly Journal of Economics 102, 857–880.

Chatterjee, S. and B. Ravikumar, 1992, A neoclassical model of seasonal fluctuations, Journal of Monetary Economics 29, 59–86.

Christiano, L.J. and M. Eichenbaum, 1990, Unit roots in real GNP: do we know, and do we care?, Carnegie-Rochester Conference Series on Public Policy 32, 7–62.

Cochrane, J.H., 1988, How big is the random walk in GNP?, Journal of Political Economy 96, 893–920.

DeLong, B.J. and L.H. Summers, 1988, How does macroeconomic policy affect output?, Brookings Papers on Economic Activity 2, 433–480.

Friedman, M. and A.J. Schwartz, 1982, Monetary trends in the United States and the United Kingdom (NBER, University of Chicago Press).

Jaeger, A., 1990, Shock persistence and the measurement of prewar output series, Economics Letters 34, 333–337.

Kuznets, S.S., 1946, National product since 1869, General Series, No. 46 (NBER, New York).

Kuznets, S.S., 1961, Capital in the American economy: its formation and financing (NBER, Princeton University Press).

Osborn, D.R. et al., 1988, Seasonality and the order of integration for consumption, Oxford Bulletin of Economics and Statistics 50, 361–377.

Osborn, D.R., 1991, The implications of periodically varying coefficients for seasonal time series processes, Journal of Econometrics 28, 373–384.

Romer, C.D., 1989, The prewar business cycle reconsidered: new estimates of gross national product, 1869–1908, Journal of Political Economy 97, 1–37.

Rudebusch, G.D., 1993, The uncertain unit root in real GNP, American Economic Review 83, 264–72.

Shaw, W.H., 1947, The value of commodity output since 1869 (NBER, New York).

Stock, J.H. and M.W. Watson, 1986, Does GNP have a unit root?, Economic Letters 22, 147–51.
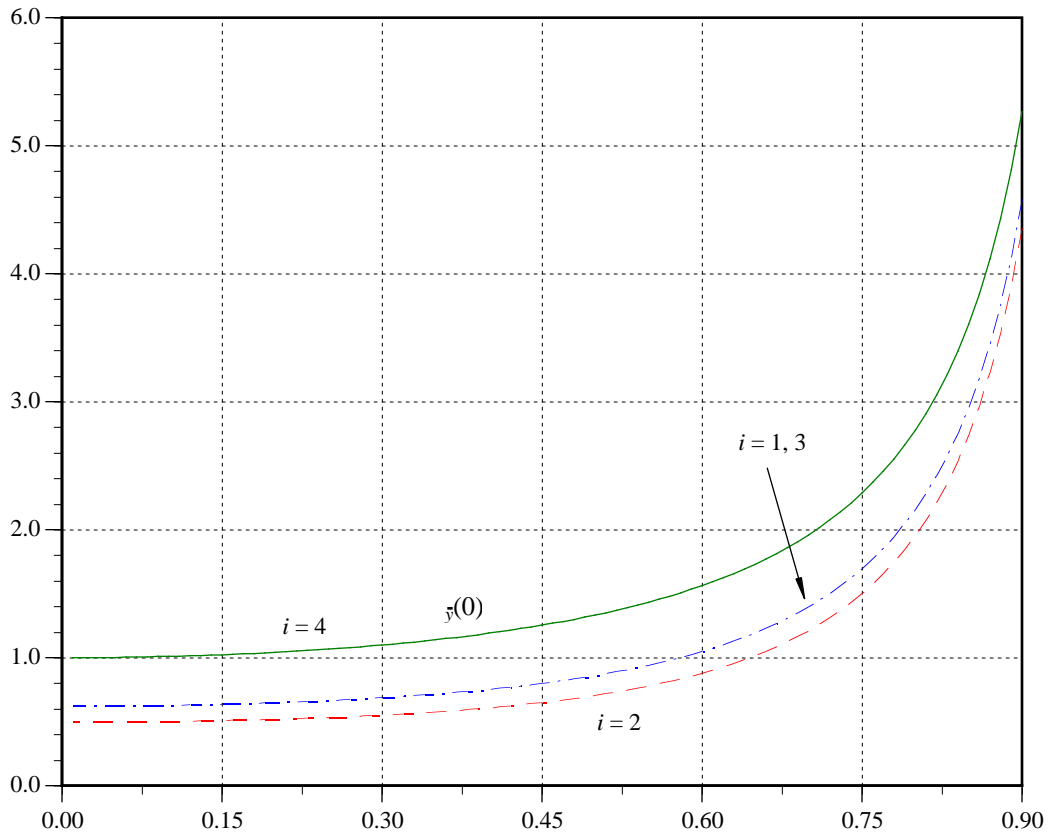
$\dot{x}, {}_i(0),\quad {}_y(0)$

Fig. 1. Variance of the original series and periodic variances of the interpolated series evaluated at various 's.
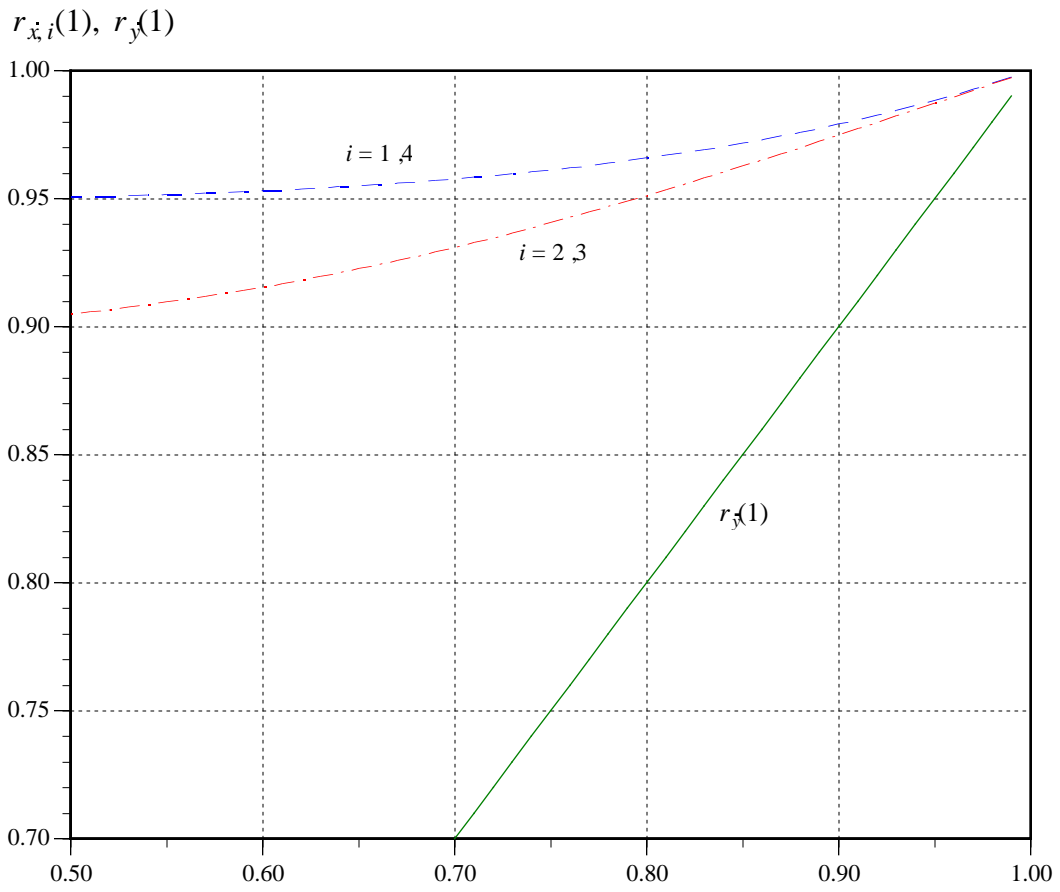
Fig. 2. First order autocorrelation of the original series and first order periodic autocorrelations of the interpolated series evaluated at various 's.
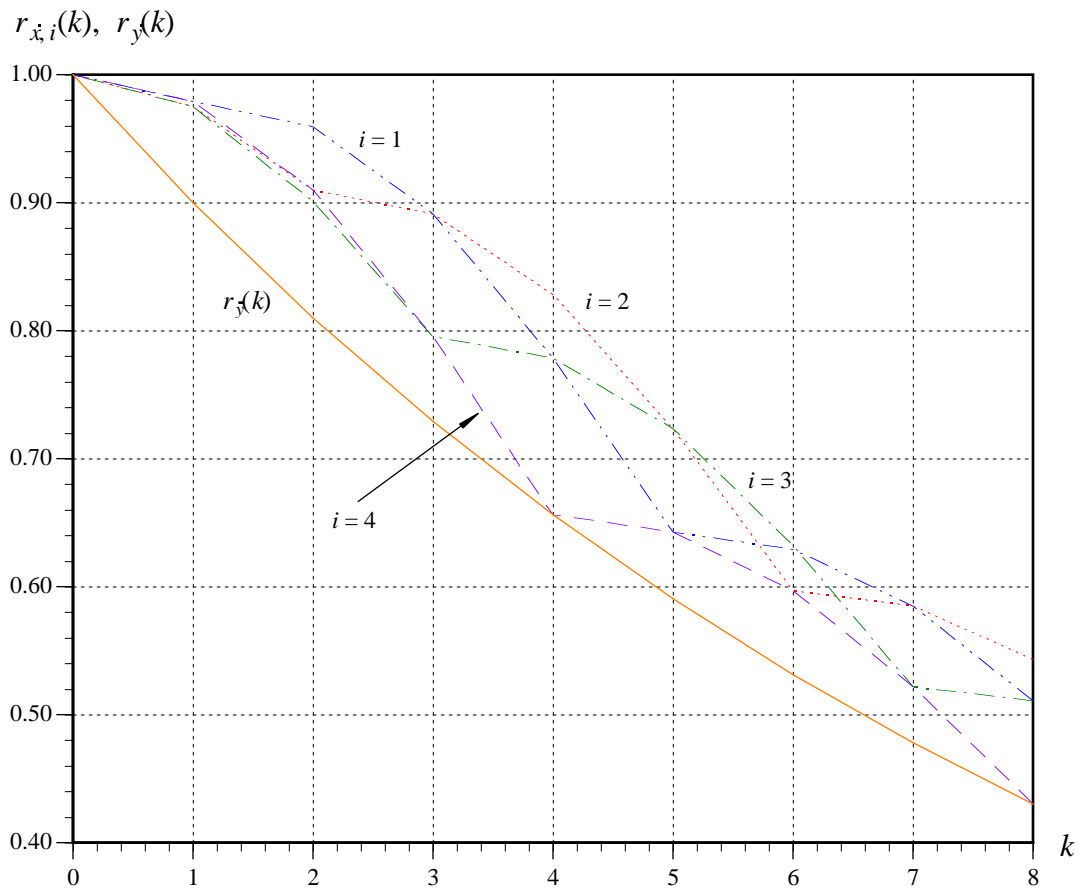
Fig. 3. Autocorrelation function of the original series and periodic autocorrelation functions of the interpolated series evaluated at various lags (k).