Note on the Correlation of First Differences of Averages in a Random Chain

Holbrook Working

Econometrica, Volume 28, Issue 4 (Oct., 1960), 916-918.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28196010%2928%3A4%3C916%3ANOTCOF%3E2.0.CO%3B2-X

Econometrica, Vol. 28, 4 (October 1960)

NOTE ON THE CORRELATION OF FIRST DIFFERENCES OF AVERAGES IN A RANDOM CHAIN

BY HOLBROOK WORKING

IN THE STUDY of serial correlations in prices series it is important to bear in mind that the use of averages can introduce correlations not present in in the original series.[1] I consider here the effect of averaging successive groups of items in a random chain, which is the simplest sort of stochastic series to which stock prices and certain commodity prices have a close resemblance.

The equation for a random chain may be written,

$$(1)\quad X_i = X_{i-1} + \delta_i \quad (i = 1,2,\ldots; E(\delta_i) = 0; \operatorname{cor}(\delta_i, \delta_{i+j}) = 0 \text{ when } j \neq 0),$$

where X_{i-1}, X_i are successive terms in the chain. In what follows I assume further, for convenience and without loss of generality, that $\operatorname{var}(\delta_i) = 1$.

Consider now a random chain that is treated as being composed of successive segments of m items each, corresponding to weeks, months or any other time intervals, in a price series. For purposes of illustration we may take $m = 3$ and write the terms of three successive segments of a random chain by derivation from the δ's as follows:[2]

i = 0	1	2	3	4	5	6	7	8	9
δ_i =	+2.0	−1.1	−0.6	+0.3	+1.3	−1.0	+0.1	+0.7	−0.3
X_i = 2.4	4.4	3.3	2.7	3.0	4.3	3.3	3.4	4.1	3.8

It is obvious that if, from this segmented random chain, we take first

[1] One example of the need is cited in the paper by Alfred Cowles elsewhere in this issue of *Econometrica*. Another example is afforded by M. G. Kendall's conclusion that wheat prices and cotton prices have behaved differently, as evidenced by a first-order serial correlation of first differences, $r_1 = +0.313$, for cotton prices as against a corresponding coefficient, $r_1 = -0.071$, for wheat prices ("The Analysis of Economic Time-Series," *Journal of the Royal Statistical Society*, CXVI (1953), pp. 15, 23). Because the cotton-price series that Kendall used consisted of monthly averages of, for the most part, daily prices, a serial correlation of about $r_1 = +0.25$ in the cotton price series was to have been expected simply as a result of the averaging process, as I show below. The wheat price series that Kendall used, on the other hand, was one that I had compiled without averaging, in order to avoid introducing the averaging effect. When this difference in constitution of the two series is taken into account there remains no clear evidence of difference in behaviour between wheat prices and cotton prices.

[2] The example is from Holbrook Working, "A Random-Difference Series for Use in the Analysis of Time Series," *Journal of the American Statistical Association*, March, 1934, taking the last figure in the second column of the table as X_0, but with all figures divided by 10 in order to have $\operatorname{var}(\delta_i) = 1$.

differences between terms correspondingly positioned in each segment, $\Delta_{i(m)} = X_i - X_{i-m}$, these first differences will have a variance,

$$\text{var}(\Delta_{i(m)}) = m \,. \tag{2}$$

But suppose that we now *average* the m terms in each segment of the chain and take first differences between the averages,

$$\overset{*}{\Delta}_{i(m)} = \frac{1}{m}(X_i + X_{i+1} \ldots + X_{i+m-1}) - \frac{1}{m}(X_{i-m} + X_{i-m+1} \ldots + X_{i-1}) \,.$$

Using the relationship in (1), as illustrated in the tabulation above, this may be written, with reversal of the order of terms in the second parenthesis above,

$$\overset{*}{\Delta}_{i(m)} = \frac{1}{m}[X_i + (X_i + \delta_{i+1}) \ldots + (X_i + \delta_{i+1} + \delta_{i+2} \ldots + \delta_{i+m-1})]$$

$$- \frac{1}{m}[(X_i - \delta_i) + (X_i - \delta_i - \delta_{i-1}) \ldots + (X_i - \delta_i - \delta_{i-1} \ldots - \delta_{i-m+1})] \,,$$

and then simplified to,

$$\begin{aligned} \overset{*}{\Delta}_{i(m)} &= \frac{1}{m}[(m-1)\,\delta_{i+1} + (m-2)\,\delta_{i+2} \ldots \\ &+ \delta_{i+m-1} + m\delta_i + (m-1)\,\delta_{i-1} \ldots + \delta_{i-m+1}] \,. \end{aligned} \tag{3}$$

Then, bearing in mind that the δ's are all mutually uncorrelated and have been assigned a variance of unity, we may derive,

$$\text{var}(\overset{*}{\Delta}_{i(m)}) = \frac{1}{m^2}[(m-1)^2 + (m-2)^2 \ldots + 1^2 + m^2 + (m-1)^2 \ldots + 1^2],$$

which reduces to,

$$\text{var}(\overset{*}{\Delta}_{i(m)}) = \frac{2m^2+1}{3m} \,. \tag{4}$$

Comparison of expression (4) with expression (2) shows that, with m only moderately large, the variance of first differences between averages over successive segments of a random chain approximates 2/3 of the variance of first differences between correspondingly positioned terms in the chain.

The covariance of successive first differences between averages may be written as, $\text{cov}(\overset{*}{\Delta}_{i(m)}, \overset{*}{\Delta}_{(i-m)\,(m)})$, which suggests that the covariance may be derived from the product of expression (3) multiplied by

$$\begin{aligned} \overset{*}{\Delta}_{(i-m)m}\ &\frac{1}{m}[(m-1)\delta_{i-m+1} + (m-2)\delta_{i-m+2} + \ldots \\ &+ \delta_{i-1} + m\delta_{i-m} + (m-1)\delta_{i-m-1} + \ldots + \delta_{i-2m+1}] \,. \end{aligned} \tag{5}$$

Inasmuch as the δ's are mutually uncorrelated, the only terms that will appear in an expression for the covariance will be those resulting from multiplication of terms in (3) and (5) that have like subscripts for the δ's. These are the terms involving $\delta_{i-m+1}, \delta_{i-m+2}, \ldots, \delta_{i-1}$. Thus we may readily derive,

$$(6)\quad \operatorname{cov}(\overset{*}{\Delta}_{i(m)}, \overset{*}{\Delta}_{(i-m)(m)}) = \frac{1}{m^2}[1(m-1) + 2(m-2)\ldots + (m-1)1] = \frac{m^2-1}{6m}.$$

Then from (4) we have,

$$(7)\quad \operatorname{cor}(\overset{*}{\Delta}_{i(m)}, \overset{*}{\Delta}_{(i-m)(m)}) = \frac{m^2-1}{2(2m^2+1)}\,.$$

From this expression it appears that even with m fairly small, the expected first-order serial correlation of first differences between averages of terms in a random chain approximates $E(r_1) = +1/4$. With $m = 5$, corresponding to weekly averages for a 5-day week, $E(r_1) = +0.235$. With $m = 2$, as it would be if monthly averages were derived by averaging prices at the ends of the first and third full weeks of each month $E(r_1) = +0.167$. I have no exact solution for the case of averages bases on the *high and low* prices of each month (or of any other time interval), but I suspect that the correlation introduced by such averaging is close to that given above for $m = 2$.

Serial correlation coefficients of higher order than the first remain zero for first differences of averages of successive groups of terms in a random chain, as may readily be shown by proceeding in a manner similar to that by which expression (6) is derived.

Food Research Institute, Stanford University