Volume Title: Monetary Trends in the United States and United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867-1975

Volume Author/Editor: Milton Friedman and Anna J. Schwartz

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-26409-2

Volume URL: http://www.nber.org/books/frie82-2

Publication Date: 1982

Chapter Title: The General Statistical Framework

Chapter Author: Milton Friedman, Anna J. Schwartz

Chapter URL: http://www.nber.org/chapters/c11403

Chapter pages in book: (p. 73 - 97)

3 The General Statistical Framework

An analysis of the long-period movements that are the subject of this volume requires freeing the data so far as possible from the effects of the shorter-term movements we call business cycles. These are brief compared to the more than a century covered by our data, averaging about four years in length for the twenty-six full reference cycles (from peak to peak) in the United States, about five years for the nineteen full reference cycles in the United Kingdom. Yet, though brief, the cyclical fluctuations are often large relative to the more gradual long-period changes. Hence temporal comparisons can be seriously distorted if the initial and terminal dates refer to different stages of the business cycle, even though the dates are separated by decades.

The method generally followed in cyclical studies of the NBER is to use the average value of a series over a full specific or reference cycle as the elementary unit of observation in studying secular movements.[1] That method gives two sets of secular observations: one constructed from trough-to-trough cycles—that is, the period between successive specific or reference cycle troughs—the other, from peak-to-peak cycles—that is, the period between successive specific or reference cycle peaks.

3.1 The Reference Phase Base as the Unit of Observation

The method we use is a slight variant: instead of a full cycle as the basic unit, we use a cycle phase. There are two kinds of phases: an expansion

1. Arthur F. Burns and Wesley C. Mitchell, in *Measuring Business Cycles* (New York: NBER, 1946), pp. 28, 141–44, use specific-cycle bases as the elementary unit of observation in studying secular movements. Moses Abramovitz uses reference-cycle bases (for sources, see chap. 11, note 1). R. C. Bird, M. J. Desai, J. J. Enzler, and P. J. Taubman find specific-cycle bases superior to reference cycle bases for accurately determining long-cycle turning points ("'Kuznets Cycles' in Growth Rates: The Meaning," *International Economic Review* 6 [May 1965]: 237–39).

phase running from a cycle trough to a cycle peak; and a contraction phase running from a cycle peak to a cycle trough. For regularly recurring cycles, the average over a cycle phase is clearly as free from cyclical effects as is the average over a full cycle composed of two successive phases. In practice, of course, neither is entirely free from cyclical effects: there is no way of eliminating by simple averaging the effects of an unusually large cyclical movement like the great contraction from 1929 to 1933.

We use the phase base rather than the cycle base because it gives us a larger number of observations on secular movements and greater flexibility. A two-period moving average of phase bases, weighted by the duration of the phases, gives the more usual cycle bases, trough-to-trough bases alternating with peak-to-peak bases.

As between specific and reference cycle phases, we have chosen to use the reference phases. The reason is that we want to compare different series with one another, and it is much easier to do so if the same chronology is used for all.

3.1.1 Phase Reference Dates

For the United States, we adopted the NBER's annual reference cycle chronology, which currently ends with a trough in 1975,[2] except that for our purposes we added turns in 1966 (peak) and 1967 (trough) that are not recognized in the official NBER chronology.[3]

For the United Kingdom, after reexamining the evidence, we revised some of the turning dates listed in the NBER reference chronology available through 1938.[4] We also extended the chronology through 1975 by examining a small collection of economic indicators as well as the turning points selected by others.[5]

2. A list of the dates ending with a trough in 1970 is given in United States Bureau of Economic Analysis, *Long-Term Economic Growth, 1860–1970* (Washington, D.C.: Government Printing Office, 1973), p. 64. G. H. Moore, *Business Cycles, Inflation, and Forecasting* (Cambridge, Mass.: Ballinger for NBER 1980), table A-1, pp. 438–39, gives monthly, quarterly, and calendar year dates ending with a trough in 1975.

Since this study was completed, the NBER has added a monthly reference cycle peak in January 1980 and a trough in July 1980, corresponding to an annual peak in 1979 and a trough in 1980.

3. See the discussion of the "pause" of 1966–67 in Solomon Fabricant, "The 'Recession' of 1969–1970," in *The Business Cycle Today*, ed. Victor Zarnowitz, NBER Fiftieth Anniversary Colloquium (New York: NBER, 1972), pp. 116–17.

4. Burns and Mitchell, *Measuring Business Cycles*, p. 79. The revisions through 1938 were to omit 1901 (trough) and 1903 (peak), and shift the 1917 peak to 1918.

5. Monthly turns in British business for the 1950s in C. Drakatos, "Leading Indicators for the British Economy," *National Institute Economic Review*, no. 24 (May 1963), p. 43, confirm our selection. Monthly turns in British growth cycles are given in Phillip A. Klein, "Postwar Growth Cycles in the United Kingdom: An Interim Report," *Explorations in Economic Research* 3, no. 1 (winter 1976): 110. The United Kingdom Central Statistical

Table 3.1 gives for each phase the initial and terminal years, the midpoint date, the duration of the phase in years, and the kind of phase (E for expansion, C for contraction)—in part 1 for the United States, in part 2 for the United Kingdom. Each phase is numbered consecutively to permit easy identification of subperiods that we use later. The phase numbered 1 for each country is the first phase for which data on the money stock are available (see chap. 4). For the United Kingdom a preceding phase numbered 0 is also shown, since we use data for some other series for that phase.

3.1.2 Computation of Phase Base

We follow standard NBER procedure in computing the phase base as a weighted average of all the observations during the phase, including both the initial and terminal turning points. The initial and terminal turning point observations are weighted one-half, the intervening observations, unity. (See equation 1 below.) Since each turning point observation is included in two successive phases, it would be given undue weight if weighted as heavily as the intermediate observations.

The inclusion of the same observation in two successive phases introduces serial correlation between successive phase bases that is considered further below.

Throughout, we use logarithms of money, income, and prices and construct phase bases by averaging logarithms, not absolute values. The reason is both economic and statistical: economically, relative changes are the main subject of interest; statistically, the logarithms are more nearly homoskedastic over time than the absolute values; that is, they have more nearly a random variability that is the same size over time. For interest rates, we construct phase bases by averaging the absolute values.

3.1.3 Weighting of Phase Bases in Statistical Computations

Because phases differ in length, the bases are averages of different numbers of observations. Hence, if the initial observations are statistically homogeneous in the sense that all are subject to the same error of measurement, the bases will not be. This feature alone would be allowed for by weighting each phase base by the number of observations from

---

Office has published monthly growth cycle turns for dates since 1958 in *Economic Trends*, no. 257 (March 1975), pp. 95–109; no. 271 (May 1976), pp. 70–80; no. 282 (April 1977), pp. 66–68. D. J. O'Dea, *Cyclical Indicators for the Postwar British Economy* (Cambridge: Cambridge University Press for NIESR, 1975), p. 39, also gives monthly turning points, 1951–72. Edward Shapiro, "Fluctuations in Prices and Output in the United Kingdom, 1921–71," *Economic Journal* 86 (December 1976): 746–58, gives quarterly turning points, 1921–38 and 1952–72. For the first period, the turns are for "classical" cycles, for the second period, for "growth" cycles.

Table 3.1 **Phase Reference Dates for the United States and the United Kingdom**

Part 1: United States

Phase Number	Phase Reference Years	Midpoint[a] Date	Duration of Phase (Years)	Kind of Phase[b]	Phase Number	Phase Reference Years	Midpoint[a] Date	Duration of Phase (Years)	Kind of Phase[b]
1	1867–69	1868.5	2	E	27	1919–20	1920.0	1	E
2	1869–70	1870.0	1	C	28	1920–21	1921.0	1	C
3	1870–73	1872.0	3	E	29	1921–23	1922.5	2	E
4	1873–78	1876.0	5	C	30	1923–24	1924.0	1	C
5	1878–82	1880.5	4	E	31	1924–26	1925.5	2	E
6	1882–85	1884.0	3	C	32	1926–27	1927.0	1	C
7	1885–87	1886.5	2	E	33	1927–29	1928.5	2	E
8	1887–88	1888.0	1	C	34	1929–32	1931.0	3	C
9	1888–90	1889.5	2	E	35	1932–37	1935.0	5	E
10	1890–91	1891.0	1	C	36	1937–38	1938.0	1	C
11	1891–92	1892.0	1	E	37	1938–44	1941.5	6	E
12	1892–94	1893.5	2	C	38	1944–46	1945.5	2	C
13	1894–95	1895.0	1	E	39	1946–48	1947.5	2	E
14	1895–96	1896.0	1	C	40	1948–49	1949.0	1	C
15	1896–99	1898.0	3	E	41	1949–53	1951.5	4	E
16	1899–1900	1900.0	1	C	42	1953–54	1954.0	1	C
17	1900–1903	1902.0	3	E	43	1954–57	1956.0	3	E
18	1903–4	1904.0	1	C	44	1957–58	1958.0	1	C
19	1904–7	1906.0	3	E	45	1958–60	1959.5	2	E
20	1907–8	1908.0	1	C	46	1960–61	1961.0	1	C
21	1908–10	1909.5	2	E	47	1961–66	1964.0	5	E
22	1910–11	1911.0	1	C	48	1966–67	1967.0	1	C
23	1911–13	1912.5	2	E	49	1967–69	1968.5	2	E

24	1913–14	1914.0	1	C	50	1969–70	1970.0	1	C
25	1914–18	1916.5	4	E	51	1970–73	1972.0	3	E
26	1918–19	1919.0	1	C	52	1973–75	1974.5	2	C

Part 2: United Kingdom

Phase Number	Phase Reference Years	Midpoint[a] Date	Duration of Phase (Years)	Kind of Phase[b]	Phase Number	Phase Reference Years	Midpoint[a] Date	Duration of Phase (Years)	Kind of Phase[b]
0	1868–74	1871.5	6	E	19	1927–28	1928.0	1	C
1	1874–79	1877.0	5	C	20	1928–29	1929.0	1	E
2	1879–83	1881.5	4	E	21	1929–32	1931.0	3	C
3	1883–86	1885.0	3	C	22	1932–37	1935.0	5	E
4	1886–90	1888.5	4	E	23	1937–38	1938.0	1	C
5	1890–93	1892.0	3	C	24	1938–44	1941.5	6	E
6	1893–1900	1897.0	7	E	25	1944–46	1945.5	2	C
7	1900–1904	1902.5	4	C	26	1946–51	1949.0	5	E
8	1904–7	1906.0	3	E	27	1951–52	1952.0	1	C
9	1907–8	1908.0	1	C	28	1952–55	1954.0	3	E
10	1908–13	1911.0	5	E	29	1955–58	1957.0	3	C
11	1913–14	1914.0	1	C	30	1958–60	1959.5	2	E
12	1914–18	1916.5	4	E	31	1960–62	1961.5	2	C
13	1918–19	1919.0	1	C	32	1962–65	1964.0	3	E
14	1919–20	1920.0	1	E	33	1965–66	1966.0	1	C
15	1920–21	1921.0	1	C	34	1966–68	1967.5	2	E
16	1921–24	1923.0	3	E	35	1968–71	1970.0	3	C
17	1924–26	1925.5	2	C	36	1971–73	1972.5	2	E
18	1926–27	1927.0	1	E	37	1973–75	1974.5	2	C

[a] Read .5 as 30 June; .0 as 1 January.
[b] E = expansion; C = contraction.

which it is computed. However, the differential weighting of the initial and terminal observations introduces an additional complication.

Let

n = duration of phase, where unit of time is interval between observations (i.e., n is number of years for annual data, number of quarters for quarterly data, etc.).

X_i $(i = 1, \ldots, n + 1)$ = observations entering into phase average, where X_1 is observation at initial turning point and X_{n+1}, at terminal turning point.

Y = phase average.

σ^2 = variance of variable indicated by subscript.

We then have

$$(1) \qquad Y = \frac{\frac{1}{2} X_1 + \sum_{i=2}^{n} X_i + \frac{1}{2} X_{n+1}}{n}$$

by the definition of the phase average. Assume that the X_i's can be regarded as statistically independent, and as all having the same variance equal to σ_X^2. We then have

$$(2) \qquad \sigma_Y^2 = \frac{1}{n^2} \left[\frac{1}{4} \sigma_X^2 + \sum_{i=2}^{n} \sigma_X^2 + \frac{1}{4} \sigma_X^2\right] = \frac{\sigma_X^2}{n^2} \left[n - \frac{1}{2}\right]$$

$$= \frac{2n - 1}{2n^2} \sigma_X^2.$$

Since the appropriate weight is inversely proportional to the variance, we take as the weight of the phase:

$$(3) \qquad w = \frac{2n^2}{2n - 1}.$$

The chief question about this derivation is the assumption of statistical independence. This assumption is less stringent than it may at first appear. Given the purpose for which we use the phase averages, namely, to average out cyclical movements, the dependence that is relevant is that which remains after we eliminate the cyclical effect. Most of the observed fairly high serial correlation in annual data, which might be taken as evidence against the assumption of independence, reflects the cyclical movement. The serial correlation between deviations from the cyclical pattern must be very much lower and, for annual data, may even be negligible. Low serial correlation will not introduce much error into the weights.

We have neglected the serial correlation because evidence on the question is lacking and because it would be different for different series, whereas it is a great convenience to use the same weights for all series.

3.1.4 Possible Difficulties with Reference Phase Bases

Three major questions arise about this procedure of using reference phase bases as the unit of observation. (1) If there are important systematic differences between the cyclical timing of a series and the reference dates, the phase bases may have a residual cyclical element. The brevity of the phases may (2) leave too much random variation and (3) introduce substantial serial correlation.

Residual Cyclical Element

The problem here is suggested by chart 3.1, which shows a regularly recurring cycle that lags in timing behind the reference cycle by one-quarter of a cycle. No problem arises for specific phase bases.

For the reference phase bases we use, however, there clearly is a problem: reference contraction bases would tend to be larger, and refer-

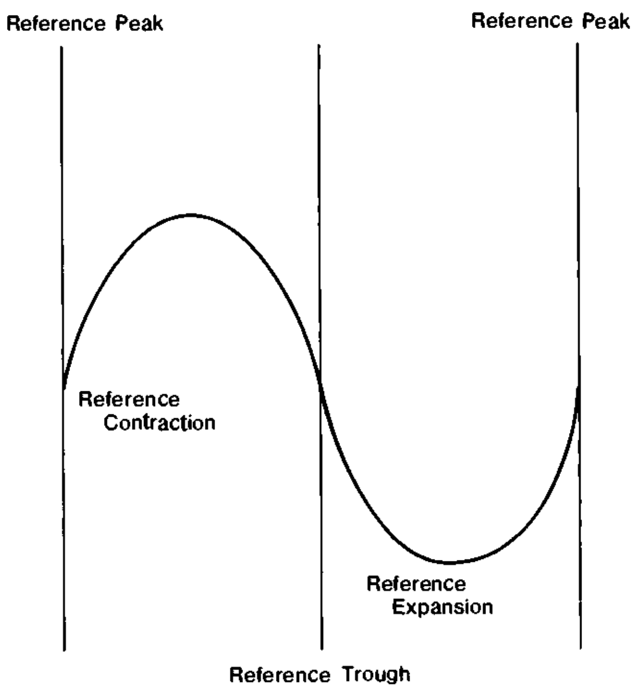

Chart 3.1 Illustration of a one-quarter specific cycle lag behind the reference cycle.

ence expansion bases smaller, than the relevant cycle-free average. Conversely, if the series leads in time, relevant contraction bases would tend to be smaller, and reference expansion bases larger, than the relevant cycle-free averages.

A lead or lag of one-quarter cycle maximizes the residual cyclical element, but clearly a lesser error of the same kind is introduced by a shorter systematic lead or lag.

One mitigating circumstance is that most of the series we use are annual and tend to be fairly synchronous with the cycle (nominal income, output, prices), so that any leads or lags are likely to be small relative to the time unit of observation.

We have made a number of tests to be sure that the residual cyclical element is negligible.

1. In the simple trend-free case pictured, each contraction base would be higher than each expansion base. In the presence of trends, this relation need not hold. To allow for trend, we have computed the average value of deviations from trend for all expansion bases and all contraction bases separately for a period chosen to begin and end with the same type of phase, which is a further protection against contamination by trend. The average values and standard deviations are shown in table 3.2. The final column gives the ratio of the mean difference to its standard error. Only one series for the United States, and only two for the United Kingdom have mean differences that are even as large as the relevant standard error, and no difference approaches statistical significance. It is interesting that the largest differences relative to their standard error are for interest rates, which are known in general to lag. As expected, the average for these series is larger for contractions than for expansions—as tends also to be true for the other series—but even for these the difference between expansion and contraction phases is clearly small enough to be neglected for any but the most refined analysis.

Note that the consistent difference in sign between the means for expansions and contractions (fifteen out of sixteen means negative for expansions and positive for contractions) is not relevant evidence of a systematic difference because the observations are not independent. All the means are calculated for precisely the same set of contractions and expansions. Given the high conformity among the series within cycles, if the expansions happen to yield negative deviations and the contractions positive deviations for one series, that is likely to be true also for others. For the same reason, the entries in the final column of table 3.2 do not provide sixteen independent bits of information, but a much smaller number.

2. A different test is to compute the rate of change between successive phases. If these are dominated by the cyclical element, they should

display a negative serial correlation—in the hypothetical case pictured, being alternatively positive and negative. Any constant percentage trend will tend to show up in the average value and not affect the serial correlation. The serial correlations obtained in this way are given in table 3.3 both for the period as a whole and for subperiods that we use in our subsequent analysis. For the United States, thirty-nine out of the fifty serial correlations are positive; for the United Kingdom, thirty-two out of forty are positive, indicating that any residual cyclical element is dwarfed by the long swings about the trends that show up in the phase bases. The negative correlations are almost all for short-term interest rates, which are highly volatile. Of these, only the postwar correlations differ significantly from zero, even at a .10 percent confidence level. So even here there is no evidence of a significant cyclical residual, only of the absence of long swings of the kind that characterize the other series.

3. On a less formal basis, we have examined the series to see whether they show many sawtooth sequences of ups and downs such as the residual cyclical element might produce. Needless to say, such sequences occasionally occur, but not frequently enough to give an impression of a significant residual cyclical, which is why we have not formalized this test.

All in all, we conclude that our phase bases do eliminate the bulk of the systematic cyclical fluctuation.

Brevity of Phases

Individual phases are often very brief. This is particularly serious for annual data. According to the annual reference dates,[6] nearly half of all phases in the United States since 1867 and more than one-quarter of all phases in the United Kingdom since 1874 have been only a year in duration (table 3.4). Phase averages for short phases do not average out much of the random movement. In addition, successive phase averages have substantial serial correlation because turning-point observations enter more than one phase.[7]

The effect of the turning-point observation is, of course, greater for annual than for monthly or quarterly data. Hence one test of the seriousness of this problem is to compare phase bases for the same series from monthly, quarterly, and annual data. For the periods for which our data permit such a comparison, the differences are always minor and mostly

6. The NBER has estimated separate sets of dates for data reported for different time units: for annual calendar year data, annual fiscal year data, quarterly data, and monthly data. See *Business Cycle Indicators*, ed. G. H. Moore, 1:670.

7. If the observations are annual, and the phase one year in length, there are no intervening observations; hence the phase average is based on values for only two years, and each of these values is used also in computing another phase average.

Table 3.2 **Average and Standard Deviations of Deviations from Trends: Expansion and Contraction Phases, for Period Beginning and Ending with Contraction**

Series	Expansion		Contraction		Difference		
	Mean	Standard Deviation	Mean	Standard Deviation	Mean (Contraction Minus Expansion)	Standard Error[c]	Ratio of Mean Difference to Standard Error
			United States (1873–1975)[a]				
1. Log nominal income	−.0151	.2705	.0274	.2055	.0425	.069	0.62
2. Log real income	.0096	.1044	.0016	.0733	−.0080	.026	−0.31
3. Log prices	−.0246	.2084	.0257	.2041	.0503	.059	0.85
4. Log money	−.0006	.1955	.0220	.1406	.0226	.049	0.46
5. Log velocity	−.0144	.2035	.0053	.1789	.0197	.055	0.36
Interest rates							
6. Call money	−.0012	.0193	.0027	.0172	.0039	.0052	0.75
7. Commercial paper	−.0019	.0177	.0037	.0170	.0056	.0049	1.14
8. Basic yield	−.0013	.0110	.0015	.0111	.0028	.0032	0.88
9. Bond rate	−.0013	.0127	.0017	.0124	.0030	.0036	0.83

			United Kingdom (1874–1975)[b]				
1. Log nominal income	−.0307	.2791	.0432	.2842	.0739	.093	0.79
2. Log real income	−.0012	.0945	−.0037	.0748	−.0025	.028	−0.09
3. Log prices	−.0295	.2580	.0469	.2654	.0764	.086	0.89
4. Log money	−.0273	.2063	.0314	.2235	.0587	.071	0.83
5. Log velocity	−.0035	.1900	.0119	.1430	.0154	.056	0.28
Interest rates							
6. Short-term rate	−.0028	.0196	.0053	.0193	.0081	.0064	1.27
7. Consol rate	−.0024	.0133	.0037	.0177	.0061	.0051	1.20

[a]Twenty-four expansions and twenty-five contractions.
[b]Eighteen expansions and nineteen contractions.
[c]Standard error is calculated as follows:

$$SE = \sqrt{\frac{\sigma_E^2}{n_E} + \frac{\sigma_C^2}{n_C}}$$

where n_E = number of expansions; n_C = number of contractions. See chapter 4 for a description of the series.

Table 3.3 **Serial Correlations of Rates of Change between Successive Phase Bases: United States, 1869–1975; United Kingdom, 1874–1975**

Series	Full Period	Peacetime Phases	Pre-World War I	Interwar	Post-World War II
		United States			
Nominal income	.42	.44	.31	.33	.57
Real income	.19	.18	.10	.24	.20
Prices	.65	.61	.51	.15	.86
Money	.52	.52	.40	.37	.80
Velocity	.30	.44	.08	.11	.74
Population	.87	.94	.72	.90	.95
Call money rate	−.13	−.20	−.36	.02	−.70
Commercial paper rate	−.28	−.35	−.63	−.24	−.58
Basic yield	.55	.55	.47	−.37	.28
Bond rate	.48	.46	.04	−.29	.23
		United Kingdom			
Nominal income	.62	.67	.48	.38	.81
Real income	.52	.08	−.45	.14	.36
Prices	.66	.69	.51	.33	.93
Money	.72	.80	.63	.33	.94
Velocity	.44	.59	−.37	.40	.86
Population	.76	.88	.67	.75	.71
Short rate	−.18	−.26	.05	−.47	−.71
Consol rate	.29	.28	.76	−.53	−.03

Note: Basic observations correlated are first differences of phase bases divided by interval between successive bases. See chapter 4 below for a description of the series. See chapter 5, note 34 for designation of wartime phases.

negligible, indicating that this effect of the overlap is not serious even for annual data.[8]

Serial Correlation

The serial correlation introduced by overlap can be readily estimated theoretically. Combining equation (10) below, which gives the covar-

8. For example, the average value of the United States money stock is as follows during the four United States reference phases, 1908–14, based on:

United States Phase Number	*Annual Reference Dates: Annual Data*	*Quarterly Reference Dates: Quarterly Data*	*Monthly Reference Dates: Monthly Data*
		(billions of dollars)	
21	12.52	12.40	12.42
22	13.72	13.77	13.79
23	15.02	15.15	15.20
24	16.06	16.08	16.08

The monthly and quarterly data are given in Friedman and Schwartz, *Monetary Statistics of the United States*, pp. 9–15, 65–66. For the annual data, see the appendix to chapter 4.

Table 3.4 **Distribution of Annual Reference Phases in the United States from 1867 to 1975 and in the United Kingdom from 1874 to 1975**

	United States			United Kingdom		
Duration (in Years)	Number of Expansions	Number of Contractions	Total	Number of Expansions	Number of Contractions	Total
1	3	20	23	3	7	10
2	11	3	14	3	5	8
3	6	2	8	4	5	9
4	3	0	3	3	1	4
4 +	3	1	4	5	1	6
Total	26	26	52	18	19	37
Mean duration	2.73	1.42	4.15	3.39	2.11	5.50

iance of successive phase bases, with equation (2) above, which gives the variance of a phase base, we have

$$r_{12} = \frac{1}{2}\frac{1}{(2n_1 - 1)^{1/2}(2n_2 - 1)^{1/2}}, \tag{4}$$

where r_{12} is the serial correlation between successive phase bases that would be produced by overlapping if the bases were otherwise statistically independent, n_1 is the duration of one of the two phases, and n_2 is the duration of the other. Some idea of the possible significance of this effect can be gained by tabulating the value of r for a number of special cases:

n_1	n_2	r_{12}
1	1	0.50
1	2	0.29
1	3	0.22
1	4	0.19
2	2	0.17

For the United States data, out of fifty-one successive pairs of phases, only four have $n_1 = n_2 = 1$, and another twenty have $n_1 = 1$, $n_2 = 2$ or $n_1 = 2$, $n_2 = 1$. Hence, while the problem of serial correlation as a result of overlap is clearly present, equally clearly it is not of major moment: a correlation coefficient of 0.29 means that only 9 percent of the variance of one phase is accounted for by its correlation with the prior phase.

3.2 Rates of Change Computed from Phase Bases

To examine in greater detail the movements in the various series over periods that are short relative to the whole period covered, though longer than the cycle phase, it is helpful to eliminate long-period trends. One way to do this is to compute trend lines and direct attention to the undulations about the trends (see charts 5.1 and 5.2). A different and frequently preferable technique is to compute rates of change between successive observations. This technique has the advantage that (1) it does not require choosing the period to cover or a specific mathematical form for the trend; (2) the observations for any one period do not depend on the far distant observations for other periods that affect fitted trends; and (3) the series can be extended either backward or forward without either recomputing or extrapolating trends. It has the disadvantages that it gives full play to measurement errors and it introduces negative serial correlation.[9]

For phase bases, the importance of measurement errors is reduced by the averaging of observations in constructing the phase average. However, as noted earlier, for annual data often only a few observations are averaged, and the overlap between successive phases introduces positive serial correlation. To reduce the measurement error, we calculate rates of change from groups of three successive phase bases. For each triplet, we calculate the slope of a least squares line, weighting each observation inversely to its variance (sec. 3.1.3). We treat this slope as the rate of change at the midpoint of the central phase of the three phases covered. It turns out that this procedure is arithmetically nearly the same as computing rates of change from overlapping cycle bases (see sec. 3.2.2).

Since a phase for the United States is on the average about 2 years in duration, and for the United Kingdom, about 2.8 years, our rates of change on the average refer to a time span of about 4 years for the United States from the midpoint of the initial phase to the midpoint of the third, and about 5.6 years for the United Kingdom. Of course the actual time span varies from date to date—from a minimum of 2 years for the United States for the triplet of phases centered on the expansion of 1919–20, to a maximum of 8.5 years for the triplet centered on the contraction of 1873–78; and from a minimum of 2 years for the United Kingdom for the triplet of phases centered on the expansion of 1919–20 or the contraction

9. Let $Y_1, Y_2, \ldots Y_n, \ldots$ be observations at time $t_1, t_2, \ldots t_n, \ldots$. Rates of change between successive observations are then given by

$$\frac{Y_2 - Y_1}{t_2 - t_1}, \frac{Y_3 - Y_2}{t_3 - t_2}, \frac{Y_4 - Y_3}{t_4 - t_3}, \ldots.$$

If the random components of the Y_i are independent of one another, then the variance of the random component of the first difference will be the sum of the variance of the random components of the two observations differenced. As the first two sample observations show, Y_2 enters positively into the first, negatively into the second; hence the random components of the two will have a negative correlation, and similarly with the second and third. However, the serial correlation will be zero under these assumptioms for observations separated by one or more other observations (e.g., first and third, etc.)

of 1927–28, to a maximum of 10.5 years for the triplet centered on the expansion of 1893–1900.

A slightly different procedure has been used extensively by Moses Abramovitz (see footnote 1, above) in his studies of long swings. He has generally computed rates of change between successive trough-to-trough cycles and also between successive peak-to-peak cycles, and then has interwoven the two sets of rates of change into a single series. His method makes each rate of change depend on four successive phases. After some experimentation, we concluded that our method gives greater sensitivity in tracing the movements in our series, with little if any loss in reliability.

3.2.1 Weights for Rates of Change

The statistical error associated with a rate of change computed in this way clearly depends on the length of the phases entering into its computation. To allow for this effect, we have weighted the rates of change in regression and similar calculations. The derivation of the weights follows.

Let

n_i = duration of phase i,
Y_i = average value of phase i,
w_i = weight given by equation (3) for phase i.

For rates of change computed from successive triplets of phases, i will take the values 1, 2, and 3. The time coordinates for the three phases measured from the midpoint of the first observation in phase 1 are

$$T_1 = \frac{n_1}{2},$$

(5)
$$T_2 = n_1 + \frac{n_2}{2},$$

$$T_3 = n_1 + n_2 + \frac{n_3}{2}.$$

Using the weights from equation (3), we have the mean time as

(6)
$$\bar{T} = \frac{\Sigma\, w_i T_i}{\Sigma\, w_i},$$

where the sums run throughout from $i = 1$ to $i = 3$. The slope of the regression fitted to the three phase averages, which we interpret as the rate of change at time T_2, is

(7)
$$b = \frac{\Sigma\, w_i Y_i (T_i - \bar{T})}{\Sigma\, w_i (T_i - \bar{T})^2}.$$

The variance of this slope is given by

$$(8)\qquad \sigma_b^2 = \frac{1}{[\Sigma\, w_i(T_i - \bar{T})^2]^2}\,[\Sigma\, w_i^2(T_i - \bar{T})^2\sigma_{Y_i}^2 + \sum_{i \neq j}\Sigma\, w_i w_j (T_i - \bar{T})(T_j - \bar{T})\sigma_{Y_iY_j}]\,.$$

By the definition of the weights

$$(9)\qquad \sigma_{Y_i}^2 = \frac{\sigma_X^2}{w_i}\,.$$

Under the assumption that the X's are independent, the covariance of the Ys differs from zero only because of common elements. Y_1 and Y_2 share the common element $\frac{1}{2}X_{n_1+1}$, Y_2 and Y_3 the common element $\frac{1}{2}X_{n_1+n_2+1}$, and Y_1 and Y_3 have no common element. It follows that

$$(10)\qquad \sigma_{Y_1Y_2} = \frac{1}{4n_1n_2}\,\sigma_X^2$$

$$(11)\qquad \sigma_{Y_2Y_3} = \frac{1}{4n_2n_3}\,\sigma_X^2$$

$$(12)\qquad \sigma_{Y_1Y_3} = 0.$$

Substituting equations (9), (10), (11), and (12) in equation (8), we have

$$(13)\qquad \sigma_b^2 = \frac{\sigma_X^2}{\Sigma\, w_i(T_i - \bar{T})^2}\,\{1 + \frac{w_2(T_2 - \bar{T})}{4n_2\,\Sigma\, w_i\,(T_i - \bar{T})^2}\,[\frac{w_1}{n_1}(T_1 - \bar{T}) + \frac{w_3}{n_3}(T_3 - \bar{T}]\}\,.$$

If $n_1 = n_3$, then $T_2 = \bar{T}$, and the term in the curly brackets is unity. For simplicity, we proceed as if this were the case, and so take as our weights for the rate of change:

$$(14)\qquad w_b = \Sigma\, w_i(T_i - \bar{T})^2\,.$$

For any given ratio $\frac{n_3}{n_1} > 1$, the error made by setting the curly bracket equal to unity will be greatest for $n_1 = n_2 = 1$, and for these values of n_1 and n_2, it will increase with n_3; and the error for any triplet of n's is the same if n_1 and n_3 are interchanged. Accordingly, we can get some idea of the maximum possible error by calculations for a few cases, as follows:

n_1	n_2	n_3	*Error in Weight* (percentage)
1	1	1	0
1	1	2	2.0
1	1	3	4.8
1	1	4	7.1
1	1	5	9.0

For our United States data, there is only one triplet for which the n's are 1, 1, 3, and no other triplet with so large a ratio of n_3 to n_1 (or n_1 to n_3). For our United Kingdom data, there is only one triplet for which the n's are 4, 1, 1, and no other triplet with so large a ratio of n_1 to n_3 (or n_3 to n_1). Hence the maximum error made for the United States by setting the curly bracket equal to unity is 5 percent, and for the United Kingdom, 7 percent. This may well be less than the error involved in assuming the original observations independent. In any event the error is negligible compared with the variation in the weights computed from equation (14), which have a range of almost 37 to 1 (i.e., the largest weight is nearly 37 times the smallest) for the United States and more than 56 to 1 for the United Kingdom.

The special case of $n_1 = n_3$, for which the calculated weights are correct, has some other features of interest. In that special case, the mean time, $\bar{T}$, equals the middate of the second phase, the calculated slope is

(15) $$b(n_1 = n_3) = \frac{Y_3 - Y_1}{n_1 + n_2},$$

that is, the difference between the first and third phase divided by the time interval between them. The weight for the slope is given by

(16) $$w_b(n_1 = n_3) = \frac{n_1^2(n_1 + n_2)^2}{2n_1 - 1}.$$

3.2.2 Relation between Rates of Change Computed from Successive Triplets of Phase Averages and from Overlapping Cycle Bases

Let Z_1 and Z_2 be two successive overlapping cycle bases. Then, by definition,

(17) $$Z_1 = \frac{n_1Y_1 + n_2Y_2}{n_1 + n_2},$$

(18) $$Z_2 = \frac{n_2Y_2 + n_3Y_3}{n_2 + n_3},$$

where, if phase 1 is an expansion phase, Z_1 will run from trough to trough, and Z_2 from peak to peak, whereas, if phase 1 is a contraction phase, Z_1 will run from peak to peak and Z_2 from trough to trough.

The time coordinates of these two cycle bases measured from the midpoint of the first observation in phase 1 are:

$$T(Z_1) = \frac{n_1 + n_2}{2} ,$$

(19)

$$T(Z_2) = n_1 + \frac{n_2 + n_3}{2} ,$$

so the difference between the two time coordinates is

(20) $$T(Z_2) - T(Z_1) = \frac{n_1}{2} + \frac{n_3}{2} .$$

Hence the rate of change between them is

(21) $$\frac{Z_2 - Z_1}{T(Z_2) - T(Z_1)} = \frac{\dfrac{n_2 Y_2 + n_3 Y_3}{n_2 + n_3} - \dfrac{n_1 Y_1 + n_2 Y_2}{n_1 + n_2}}{\dfrac{n_1 + n_3}{2}} ,$$

which can be reduced to

(22) $$2 \frac{-n_1(n_2 + n_3)Y_1 + n_2(n_1 - n_3)Y_2 + n_3(n_1 + n_2)Y_3}{(n_1 + n_2)(n_2 + n_3)(n_1 + n_3)} .$$

Consider now the slope of a least squares line fitted to Y_1, Y_2, and Y_3, where the phase averages are weighted by n_i instead of w_i, as given by equation (3). This would be the correct weight if successive phase averages had no items in common but were constructed as a straight average of the relevant number of observations rather than by giving half-weight to the initial and terminal observations.

The slope of such a regression would be given by

(23) $$b' = \frac{\Sigma\, n_i Y_i (T_i - \bar{T}')}{\Sigma\, n_i (T_i - \bar{T}')^2} ,$$

where T_i are given by equation (5), and

(24) $$\bar{T}' = \frac{\Sigma\, n_i T_i}{\Sigma\, n_i} = \frac{n_1 + n_2 + n_3}{2} ,$$

so that

$$T_1 - \bar{T}' = -\frac{n_2 + n_3}{2} ,$$

$$T_2 - \bar{T}' = \frac{n_1 - n_3}{2}, \tag{25}$$

$$T_3 - \bar{T}' = \frac{n_1 + n_2}{2}.$$

Substituting equations (25) in equation (23) and simplifying gives

$$b' = 2\,\frac{-n_1(n_2+n_3)Y_1 + n_2(n_1-n_3)Y_2 + n_3(n_1+n_2)Y_3}{(n_1+n_2)(n_2+n_3)(n_1+n_3)}; \tag{26}$$

which is identical with equation (22).

It follows that the rates of change we compute would be identical with the rates of change computed from two overlapping cycle bases if we used n_i instead of w_i as weights.

3.2.3 Possible Difficulties with Rates of Change Computed from Phase Bases

The same difficulties that were considered for the phase bases are relevant for the rates of change. It turns out that a residual cyclical element is a less serious problem, while spurious serial correlation may be a more serious one.

Residual Cyclical Element

The rates of change can be divided into two classes: those computed from a contraction, an expansion, and a contraction (CEC rates) and those computed from an expansion, a contraction, and an expansion (ECE rates). Even if the original phases have a residual cyclical element, the rates of change should reflect this residual cyclical element in greatly diluted form. For example, if $n_1 = n_3$, then the rate of change depends only on the two end phases, both of which are the same kind and hence subject to the same bias, if any. That is not completely true when $n_1 \neq n_3$, but presumably is largely so.

However, to make sure that this conclusion is correct, we compared CEC and ECE rates for a considerable number of our series. For eighteen United States series and sixteen United Kingdom series, we classified the rates according to whether they were above (+) or below (−) the mean. The results, for all series combined, are as shown in table 3.5

The chi-square value for the United States contingency table is .004, for the United Kingdom table, .35. The United States value would be exceeded by chance over 95 percent of the time, the United Kingdom value, over 40 percent of the time.

Similar tables for smaller groups of series yielded the same result.

Table 3.5 **Number of Rates of Change above or below the Mean, by Kind of Rate, for Selected Series**

Kind of Rate	United States			United Kingdom		
	Number of Rates			Number of Rates		
	+	−	Total	+	−	Total
CEC	221	229	450	126	162	288
ECE	229	213	442	133	147	280
Total	450	442	892	259	309	568

In addition, we calculated t tests for the difference between the means of the CEC and ECE rates for individual series. These were uniformly not statistically significant.

Accordingly, we have concluded that our rates of change are not affected by the kind of triplet of phases from which they are computed.

Spurious Serial Correlation

The possibility of spurious serial correlation arises from two sources: (1) the serial correlation between successive phase bases arising from the turning-point observation common to them; (2) the phase bases common to different rates of change. Two consecutive rates of change are based on triplets of phases that have two phases in common; two nonconsecutive rates of change separated by one rate are based on triplets that have one phase in common. Only every third rate is based on nonoverlapping triplets, and even that is contaminated by effect 1 arising from the turning point common to the terminal phase of the first triplet and the initial phase of the second triplet.

A full examination of the size of these spurious correlations would be inordinately complex. However, we can gain an impression of their possible magnitude by considering the special case examined earlier, that in which $n_1 = n_3$. For that case, it is reasonably straightforward to derive mathematically the spurious serial correlation.

We shall consider separately three cases: (1) consecutive rates of change; (2) rates of change separated by one rate; (3) rates of change separated by two rates. More distant rates should be independent of any spurious correlation arising from common elements. We shall then (4) present some empirical evidence on the actual serial correlations in our computed rates of change.

1. *Consecutive rates of change.* Let

b_{13} = rate of change computed from phases 1, 2, and 3,
b_{24} = rate of change computed from phases 2, 3, and 4.

Assume that

$$n_1 = n_3 \,, \tag{27}$$

$$n_2 = n_4 \,, \tag{28}$$

so that from equation (15)

$$b_{13} = \frac{Y_3 - Y_1}{n_1 + n_2} \,, \tag{29}$$

$$b_{24} = \frac{Y_4 - Y_2}{n_1 + n_2} \,. \tag{30}$$

We have that

$$r_{b_{13}b_{24}} = \frac{Eb'_{13}b'_{24}}{\sigma_{b_{13}}\sigma_{b_{24}}} \,, \tag{31}$$

where primes represent deviations of the slopes from their mean values, and E stands for expected value. The standard deviations are given by the square root of the reciprocal of w_b times σ_x or, by equation (16), by

$$\sigma_{b_{13}} = \frac{\sqrt{2n_1 - 1}}{n_1(n_1 + n_2)}\,\sigma_x \,, \tag{32}$$

$$\sigma_{b_{24}} = \frac{\sqrt{2n_2 - 1}}{n_2(n_2 + n_3)}\,\sigma_x \,. \tag{33}$$

To estimate the covariance of the slope, multiply b'_{13} by b'_{24}, and take expected values. This gives

$$Eb'_{13}b'_{24} = \frac{EY'_3Y'_4 - EY'_3Y'_2 - EY'_1Y'_4 + EY'_1Y'_2}{(n_1 + n_2)(n_2 + n_3)} \,. \tag{34}$$

Equations (10) and (11) plus the counterpart of equation (10) for phases 3 and 4 and of equation (12) for phases 1 and 4 give the covariances. Substituting, and using equations (27) and (28), we get

$$Eb'_{13}b'_{24} = \frac{\dfrac{1}{4n_1n_2}}{(n_1 + n_2)(n_2 + n_3)}\;\sigma_x^2 \,. \tag{35}$$

Substituting equations (32), (33), and (35) into equation (31), we have

$$r_{b_{13}b_{24}} = \frac{1}{4\sqrt{2n_1 - 1}\,\sqrt{2n_2 - 1}} \,. \tag{36}$$

This correlation reaches its maximum value for $n_1 = 1$, $n_2 = 1$, or, by equations (27) and (28), for four successive phases all one year in length. For that extreme case, of which there is none in either our United States or United Kingdom record, the serial correlation is 0.25. For $n_1 = 1$ and $n_2 = 2$, or four phases lasting 1, 2, 1, 2 years respectively, the serial correlation is 0.14. There are three such quadruplets of phases in the United States record, none in the United Kingdom record. All other quadruplets involve more uneven numbers. We conclude that this source of serial correlation can readily be neglected.

2. *Nonconsecutive rates of change separated by one rate.* For our special case, for which

$$(n_1 = n_3 = n_5) \,, \tag{37}$$

the slopes are given by

$$b_{13} = \frac{Y_3 - Y_1}{n_1 + n_2} \,, \tag{29}$$

and

$$b_{35} = \frac{Y_5 - Y_3}{n_1 + n_4} \,. \tag{38}$$

If the initial X's are all statistically independent, then so are Y_1, Y_3, and Y_5, since they contain no common elements. The slopes are not, however, statistically independent, since Y_3 enters positively into one and negatively into the other. The correlation between b_{13} and b_{35} will be precisely the same as that between successive first differences of a series of statistically independent observations, which is well known to be

(39)[10] $$r_{b_{13}b_{35}} = -\frac{1}{2} \,.$$

10. A formal proof for this special case is readily given. We have

$$Eb'_{13}b'_{35} = \frac{E(Y'_3 - Y'_1)(Y'_5 - Y'_3)}{(n_1 + n_2)(n_1 + n_4)}$$

(a) $$= \frac{-EY'^2_3}{(n_1 + n_2)(n_1 + n_4)} = \frac{-\left(\dfrac{2n_1 - 1}{2n_1^2}\right)}{(n_1 + n_2)(n_1 + n_4)} \sigma_x^2$$

by equations (2), (12), and (37). The standard deviations are given by equation (32), and a comparable equation for b_{35}, so

(b) $$r_{b_{13}b_{35}} = \frac{-\dfrac{-(2n_1 - 1)}{2n_1^2(n_1 + n_2)(n_1 + n_4)} \sigma_x^2}{\dfrac{\sqrt{2n_1 - 1}}{n_1(n_1 + n_2)} \dfrac{\sqrt{2n_1 - 1}}{n_1(n_1 + n_4)} \sigma_x^2} = -\frac{1}{2} \cdot$$

A serial correlation so large in absolute value clearly presents a potentially troublesome problem. While it does not bias correlation results between different series, it does introduce serial correlation of residuals and affects the validity of tests of statistical significance. How serious such a serial correlation is in practice depends on the size of the random element in the variability compared with the systematic element. The spurious negative correlation introduced into the rates of change contaminates our results less than a similar correlation would for first differences between temporally consecutive items for a number of reasons. First, it affects only alternate items. Second, the items affected are separated more widely in time—the average interval between two alternate rates of change is four years for the United States and five and a half years for the United Kingdom, which raises the importance of the systematic component relative to the random component.

3. *Nonconsecutive rates of change separated by two rates*. The relevant slopes for this case are given by

$$b_{13} = \frac{Y_3 - Y_1}{n_1 + n_2}, \tag{29}$$

and

$$b_{46} = \frac{Y_6 - Y_4}{n_4 + n_5}. \tag{40}$$

The only source of serial correlation is the turning-point observation common to Y_3 and Y_4. We have that

$$Eb'_{13}b'_{46} = \frac{EY'_3Y'_6 - EY'_3Y'_4 - EY'_1Y'_6 + EY'_1Y'_4}{(n_1 + n_2)(n_4 + n_5)}. \tag{41}$$

Of the terms in the numerator, all are zero (on the assumption that the original X's are statistically independent) except for $EY'_3Y'_4$ which, by equation (10), has the value $\frac{1}{4n_3n_4}\sigma_x^2$. It follows that, using the relevant counterparts of equations (31), (32), and (33) and the assumption $n_1 = n_3$,

$$r_{b_{13}b_{46}} = \frac{-\frac{1}{4n_3n_4(n_3+n_2)(n_4+n_5)}\sigma_x^2}{\frac{\sqrt{2n_3-1}}{n_3(n_3+n_2)}\frac{\sqrt{2n_4-1}}{n_4(n_4+n_5)}\sigma_x^2}$$

$$= -\frac{1}{4\sqrt{2n_3-1}\sqrt{2n_4-1}}. \tag{42}$$

Table 3.6 **Serial Correlations* of Rates of Change for the United States and the United Kingdom**

	United States				United Kingdom			
		Rates Separated by				Rates Separated by		
Series	Consecutive Rates 3–50[a]	One 3–49	Two 3–48	Three 3–47	Consecutive Rates 2–35	One 2–34	Two 2–33	Three 2–32
Nominal income	.585	−.019	−.237	−.054	.668	.101	−.161	−.078
Real income	.356	−.384	−.481	−.139	.538	.034	−.200	−.185
Prices	.747	.276	−.022	−.032	.675	.140	−.180	−.220
Money	.652	.022	−.365	−.357	.735	.191	−.208	−.357
Velocity	.489	−.035	.020	.294	.632	.125	−.019	.063
Population	.915	.764	.605	.428	.830	.548	.333	.221
Short-term rate	.406	−.155	−.263	−.093	.210	−.007	.150	.057
Short-term rate	.450	−.024	−.099	.101				
Long-term rate	.744	.427	.336	.398	.744	.599	.413	.269
Long-term rate	.740	.425	.314	.332				
Number of observations	48	47	46	45	34	33	32	31
*Value of correlation coefficient that would be exceeded less than one time in 20:	.28	.28	.28	.29	.33	.33	.34	.34

[a]Initial phase numbers.

This correlation is clearly a maximum when $n_3 = n_4 = 1$, when it has the value -0.25. This correlation is opposite in sign but numerically equal to that for successive rates of change involving the same duration of the phases with the overlapping turning point. As for that case, it seems clear that this source of serial correlation is so trivial that it can readily be neglected.

4. *Some empirical evidence on serial correlation.* Table 3.6 gives serial correlations for rates of change of our main United States and United Kingdom annual series. If the spurious correlation were the only source of correlation, we would observe positive correlations for consecutive rates, all less than .25 and generally much less so; for rates separated by one, we would observe negative correlations, all around $-.5$; for rates separated by two, we would observe negative correlations, all less than .25 in absolute value and generally much less so; for rates separated by three, we would observe serial correlations of zero. Clearly, the serial correlations in table 3.6 do not correspond to this picture, either in detail or in general, except for the substantial number of insignificant correlations for rates separated by three. Any spurious correlation has apparently been eliminated by the systematic relations between the rates of change, a relation that reflects the existence of the long swings that have by now been so widely recognized, and that we analyze in chapter 11.

It is worth emphasizing that serial correlations do not introduce any bias into correlations among statistically independent series. They do affect the precise validity of tests of statistical significance, and they do affect the serial independence of residuals. On the whole, both the theoretical analysis and the empirical evidence in table 3.6 justify the conclusion that the "noise" introduced by the serial correlations is sufficiently small relative to the systematic variation we are trying to describe that it can for the most part be neglected.