Volume Title: Monetary Statistics of the United States: Estimates, Sources, Methods

Volume Author/Editor: Milton Friedman and Anna Jacobson Schwartz

Volume Publisher: NBER

Volume ISBN: 0-870-14210-0

Volume URL: http://www.nber.org/books/frie70-1

Publication Date: 1970

Chapter Title: General Features of Our Estimates

Chapter Author: Milton Friedman, Anna Jacobson Schwartz

Chapter URL: http://www.nber.org/chapters/c5290

Chapter pages in book: (p. 315 - 333)

10

GENERAL FEATURES OF OUR ESTIMATES

1. Choice of Date to Begin Monthly Estimates

FOR BOTH ECONOMIC and statistical reasons we chose to begin our monthly estimates with May 1907. It seemed desirable to have monthly figures for as long a period as possible before the Federal Reserve System was in operation. However, the data at midyear and call dates become less and less satisfactory the further one pushes back before April 28, 1909, the date of the first comprehensive balance sheet of nonnational banks by states, compiled by the National Monetary Commission.[1] By starting with 1907, we avoided placing much reliance on the earlier unreliable data, yet were able to cover two pre-World War I reference cycles, making possible at least some comparison of the detailed cyclical pattern of the money stock before and after the inauguration of the Federal Reserve System.

To avoid duplicating the current estimates that have been published in the *Federal Reserve Bulletin* for monthly dates beginning December 1942, our estimates end in 1946 for mutual savings bank deposits,

[1] Oddly enough, for these early years data for monthly interpolation, in the weekly city clearinghouse returns for that sample of banks, are superior to the state returns available at less regular intervals for larger samples. The balance sheet data reported by many states before 1909 present difficulties that weaken the reliability of call date bench marks compared to later years. These defects are largely corrected in the revised annual series of the Federal Reserve for the United States as a whole in *All-Bank Statistics, United States, 1896–1955*, Board of Governors of the Federal Reserve System, Washington, D.C., 1959. But the Federal Reserve series still does not overcome a basic shortcoming of the individual state data: the lack of consistency in dating of annual returns in the earlier years. The aggregate is a sum of returns dated "on or about June 30." The spread of dating about June 30 widens, and the number of states departing from the most common date increases, the earlier the year.

1945 for commercial bank deposits, and 1942 for currency held by the public. The corresponding Federal Reserve monthly series, seasonally adjusted, have been used to continue our series. Although our series are not constructed in the same way as those of the Federal Reserve, the two seem reasonably consistent during the period in which they overlap. The levels are similar at the dates when the shift from our estimates to the Federal Reserve series is made. The series therefore give a continuous monthly record of changes in the money stock from May 1907 to date.

2. Sources of the Data

We took most of the raw data, reported annually, at call dates, monthly, or weekly, from printed sources, principally the annual reports of the Comptroller of the Currency, the Federal Reserve Board, and individual state banking departments; the *Federal Reserve Bulletin; Banking and Monetary Statistics; All-Bank Statistics;* call reports of national, member, and insured banks; clearinghouse returns of member (and certain nonmember) banks shown in various sources including A. P. Andrew's "Statistics of Banks and Banking in the United States" (Part II of *Statistics for the United States, 1867–1909*) and the *Commercial and Financial Chronicle.* We also used unpublished data provided by various state banking departments.

The bank balance sheet information in these sources varies in pertinent detail. In general, federal reports are superior to state reports, although some state reports are more satisfactory sources of nonnational bank information for some years than are the Comptroller of the Currency's compilations.

Until comparatively recent years bank vault cash, demand deposits adjusted, and time deposits adjusted were not itemized as such on most bank balance sheets. They must be compiled by addition (or subtraction) of components listed separately. Special problems exist where no breakdowns are given: the asset side of the balance sheet may show vault cash combined with amounts due from banks or items in the process of collection; there may be no separate entry of items constituting "float"; the liability side may show demand and time deposits as a total or combined with amounts due to banks. Usually U.S. government deposits are not classified apart from those of the public. Our treatment of these problems is discussed in subsequent chapters. Our object was to make our final series as comprehensive as their defined content and to rid them of all extraneous elements. Some estimation was required to achieve this result.

Our annual series after 1896, except for mutual savings bank deposits and vault cash, are Federal Reserve constructions representing the most complete information available. These data are based on an enumeration of banks. It is conceivable that some institutions that should be considered banks may have been overlooked; it is most unlikely that any nonbanking institutions were included. Despite the vagaries of bank reports, deposits of the enumerated banks are not likely to be significantly over- or underestimated when data for other balance sheet items are simultaneously compiled or estimated.

Our call date figures, which conform in definition to the annual series, represent a considerable, though varying, proportion of the total universe. Some of the monthly data, however, differ in content from the call date series for which they served as interpolators. The use of these conceptually different data to estimate inter-call date movements does not introduce any appreciable error since (a) we first ascertained that these related data correlate well with our call date figures; and (b) the levels of our bench-mark series are accurate within the limits of our performance.

3. Classification of Banks

The basic sources of data determine the classification of banks that we have used. Prior to 1914 the obvious classification is into national and nonnational banks. Data for the national banks are excellent, homogeneous, and readily available from the reports of the Comptroller of the Currency. Data for nonnational banks must be gleaned from reports of state supervisory agencies or from the irregular and variable reports made to the Comptroller in response to his request. This division between national and nonnational banks is relevant for our deposit estimates and affects not only them, but also our estimates of bank vault cash and hence of currency in the hands of the public.

For the period since 1914 an additional classification is possible—into banks that are members of the Federal Reserve System and banks that are nonmembers. Since member banks include all national banks plus some other banks as well, we have preferred to use this classification wherever possible. However, satisfactory data are available for member banks only from June 1919 on. Hence we have been able to use only the national-nonnational division up to that date. We have carried our estimates for this division beyond that date to 1922 because for 1919–22 national banks reported at two more call dates each year than member banks did.

Even after 1922, however, we have had to use both classifications to some extent, since data on nonmember banks are not directly available. Each state receives reports from banks under its jurisdiction, and these reports usually do not separate member from nonmember banks. As a result, we generally derive estimates for nonmember banks as the difference between totals for nonnational banks and corresponding figures for state member banks taken from Federal Reserve reports. This frequently raises problems of comparability, and care was always required to assure that our estimation was carried through in such a way as to use the independent information for state member banks, since it is easy and tempting to follow procedures that in effect discard this information.

The nonnational banks are further classified into various groups in various states: commercial banks, loan and trust companies, stock savings banks, mutual savings banks; and incorporated banks and private banks; and so on. It is with these classifications, often varying in number and meaning from state to state, that we have had to work in processing data for individual states. In combining data for different states, however, we have tried to keep distinct only two categories, commercial banks and mutual savings banks (see definitions in section 7 of this chapter). In addition, of course, we have kept separate savings and loan associations, which are not regarded as banks.

4. Dating and Seasonal Adjustment

Our final series were not constructed in one continuous operation from the initial to the terminal dates. In accordance with the data available different procedures were followed in different time periods. Because a variety of devices was adopted—as will be seen from the detailed chapters—it is extremely difficult to summarize our methods adequately. We discuss in this and the next section a few aspects of our procedure that have general relevance and that supplement the brief description of our methods in Chapter 1.

Dating

Our raw data include time series in different time units: annual, call date, monthly, and weekly series. Most of the annual figures are for "June" dates. These are the dates of the figures published in the *Annual Report* of the Comptroller of the Currency. In a few instances these fall in April, May, or July. In all other cases they lie somewhere in June, usually toward the end of the month. The annual and call date series refer to single dates. Some of our monthly and weekly series are averages of daily figures, while others refer to single dates.

Particular attention was paid to the dating of the series utilized in making the monthly estimates of vault cash and currency because there is a consistent daily as well as monthly seasonal characterizing these data. Elimination of the double seasonal affecting these items (see the next part of this section) made it possible to adjust to a common set of dates all reported figures, whatever their original dating, and to use them in constructing the final series.

Intraweekly variability presented no problem for deposit data, since the size of this daily movement is small relative to the monthly seasonal.[2] The questions with respect to dating were: (a) How distant from the most common date at midyear or at call dates could the report of a group of banks be dated and still be considered eligible for inclusion in the total? (b) What adjustments should be made if the dating of a potential monthly interpolator or monthly component differed from the call date or annual series to be interpolated? (c) What adjustment should be made if the dating of potential monthly interpolators or monthly components changed from time to time? The specific answers to these problems are detailed in the succeeding chapters. Here we simply note that the prevailing end-of-month dating of the monthly interpolators led us to produce estimates dated in the same way, though beginning 1947 (for currency and commercial bank deposits) we have shifted to the current Federal Reserve series of monthly averages of daily figures and have centered at midmonth other Federal Reserve data available only for single dates near the end of the month.

[2] See Chapter 13, footnotes 6 and 7.

Seasonal Adjustment

The daily and monthly variability of the vault cash and currency data has been mentioned as a compelling reason for seasonal adjustment of these items before their employment in making estimates. Our views on interpolation procedures, which are explained briefly below, also emphasize the importance of deseasonalizing data before interpolation. We therefore examined all our data—on vault cash, currency, and deposits—for evidence of seasonality and eliminated it whenever this seemed necessary.

We measured seasonal variations principally by the method of averaging ratios to a twelve-month moving total. Although our preference is for long seasonal periods, which permit random movements of a series to cancel one another, some of the series from which we eliminated the seasonal were available for brief periods only.

Since many of our series were in call date form, we had to adapt monthly seasonal analysis to the requirements of data reported irregularly during the course of a year.

For the vault cash and currency series, characterized by daily as well as monthly movements, we estimated simultaneously daily and monthly seasonal factors. Our procedure involved the following steps:

(1) The data to be used in estimating the seasonals were expressed as relative deviations from a trend.

(2) The deviations from the trend were cross-classified by the day of the week and the month of the year.

(3) The entries in each day-month cell were averaged.

(4) A multiple regression was computed from these averages, each weighted by the corresponding number of observations, relating the average value to a set of dummy variables representing the days of the week and the months of the year (i.e., a variable which is unity for Monday and zero for other days, another which is unity for Tuesday and zero for other days, and so on; a variable which is unity for January and zero for other months, another which is unity for February and zero for other months, and so on).

(5) The daily seasonal indexes are then given by the differences between the regression coefficients for the separate days and their average value.

(6) The monthly seasonal indexes are given by the differences between the regression coefficients for the separate months of the year and their average value.

The monthly and daily indexes can be combined to yield factors for January for each day of the week, for February for each day of the week, and so on (see the appendix to Chapter 12).

5. Interpolation Procedures

Interpolation to fill in gaps plays an important role in most estimates of comprehensive economic data. It plays a particularly important role in our monthly estimates because of the lack of uniformity of dates for which basic data are available: call dates that are deliberately varied from year to year and that may differ between national and other banks, between nonnational banks in one state and nonnational banks in another, and even between different groups of nonnational banks in the same state (e.g., loan and trust companies and other commercial banks; mutual savings banks and commercial banks). We thus had the choice of discarding most of the available data or of using interpolation devices, not only to fill in gaps, but also to shift dates so as to obtain components that could be added.

The problem of interpolation was so pervasive that a fresh analysis of the general problem of interpolation was undertaken by one of us and the development of a new method of interpolation emerged as a by-product. We refer the reader to the separate publication reporting that analysis [3] for a systematic discussion of interpolation procedures. We limit ourselves here to describing and naming the specific methods that we used, in order to be able to refer to them by name in subsequent chapters and so avoid tedious repetition.

The general idea of interpolation is that the value of a series, say X, is required for dates for which there are no direct observations but that are intermediate between dates for which there are observations. A subclass of interpolation procedures, which cover all those we actually used, is comprised by procedures that use information on X for only one preceding and one subsequent date (say dates t_0 and t_2) to estimate a value for a date for which a value is required, say t_1.[4]

[3] Milton Friedman, *The Interpolation of Time Series by Related Series*, New York, NBER, Technical Paper 16, 1962 (reprinted from *Journal of the American Statistical Association*, Dec. 1962, pp. 727–757).

[4] There may of course be several different dates for which a value is required between a particular pair of known values at t_0 and t_2.

Method L

The interpolated value of X at time t_1, say X_1^*, may be estimated solely from its values at times t_0 and t_2, say X_0 and X_2. This is the method of *Linear Interpolation,* which we shall call *Method L.*[5] The general formula is

$$X_1^* = X_0 + \frac{t_1 - t_0}{t_2 - t_0}(X_2 - X_0) \tag{1}$$

$$= \frac{t_2 - t_1}{t_2 - t_0} X_0 + \frac{t_1 - t_0}{t_2 - t_0} X_2 \tag{2}$$

or a weighted average of X_0 and X_2. As long as only the two known values of X are used, this method is perfectly general—a wide variety of particular forms being generated by appropriate transformations of the observations. For example, if X is in dollars, this is linear interpolation of dollar amounts; if X is the logarithm of dollar amounts, this is linear interpolation of the logarithms or logarithmic interpolation of the dollar amounts, and so on for other transformations. We shall refer to X_1^* sometimes as the interpolated value, sometimes as the *trend value* of X, where by *trend* we mean no more than a straight line connecting two successive values of X.

For many actual economic time series we may have information about the movement of X between t_0 and t_2 that is provided not by X_0 and X_2 only but by the date t_1. This information comes from the knowledge that series X is subject to a recurrent seasonal movement that we may be able to estimate from known data for other periods or other series. In such cases, Method L alone is inefficient. The better procedure is to eliminate the known or estimated seasonal and then apply Method L to the deseasonalized values of X_0 and X_2. If the value desired is a seasonally uncorrected value, the seasonal should then be added back.[6]

We followed this procedure throughout, deseasonalizing the original data, as noted in section 4 above, before applying any interpolation method. We accepted a deseasonalized series as our end-product and, therefore, did not add a seasonal back in. Unless otherwise stated, therefore, all series used in interpolation procedures will be assumed to be free from seasonal.

[5] In Technical Paper 16, this method was called Method M_0. We have changed the designation here for mnemonic reasons.
[6] See Technical Paper 16, section IV, pp. 22–23.

Method R

Generally we should like to do better than linear interpolation. In order to do so, we must have additional information that can tell us something about how X changes relative to its trend between t_0 and t_2. The most obvious source of such information is the movement of another series that we have reason to believe is related to X—in the sense that the two move together—and the value of which is known at t_0, t_1, and t_2. Let Y designate a hypothetical example of such a series. (We shall discuss later our method of choosing a related series.) Clearly, the more closely the movements of Y are related to the movements of X, the more useful Y will be for interpolation. If the relation is small or negligible, the use of a related series may simply introduce a random component without improving the accuracy of the interpolated value. Under such circumstances, we have preferred to use Method L.

Given a related series, we have five known observations: X_0, X_2, Y_0, Y_1, Y_2. We want to use the final three to improve the estimate by Method L of X_1. Let Y_1^* be a trend value for Y at time t_1, comparable to X_1^* for X. Let

$$u = \frac{X_1 - X_1^*}{X_1^*}, \tag{3}$$

$$v = \frac{Y_1 - Y_1^*}{Y_1^*}, \tag{4}$$

i.e., the percentage deviations of X and Y from their trend values at time t_1.[7] Then a special case of interpolation by related series, and the only one we used, is to estimate u from v. If we knew the correlation of u and v (ρ_{uv}) and the standard deviations of u and v (σ_u and σ_v), and if we could assume that the means of u and v are both zero (which is what our advance deseasonalization is intended to help assure), the correct estimate would clearly be as follows:

$$\text{Estimate of } u = \rho_{uv} \frac{\sigma_u}{\sigma_v} \cdot v = \beta \cdot v. \tag{5}$$

Needless to say, we cannot know ρ, or the σ's. We have, however, made empirical estimates of them, say r_{uv}, s_u, and s_v, and, maintaining

[7] In some ways it is more elegant to define u and v as simply equal to $X_1 - X_1^*$ and $Y_1 - Y_1^*$, which would make them percentage deviations if X and Y were logarithms of the original observations. The formulas above mix logarithmic and arithmetic operations. However, there is no logical fallacy involved in mixing the operations, the quantitative difference is negligible for most of our series, and the procedure described is slightly simpler computationally. At any rate, it is the procedure we used. See Technical Paper 16, pp. 4–5.

the assumption that the means of u and v are zero, have formed the estimate

$$u^* = r_{uv} \frac{s_u}{s_v} \cdot v = b \cdot v, \tag{6}$$

which gives us an estimate of X_1,

$$X_1^{**} = X_1^*(1 + u^*). \tag{7}$$

Note that Method L can be regarded as a special case of Method R for $b = 0$.

We may distinguish three special cases according to the estimates used for b and for v.

(i) *Method* R_1.[8] We find (or believe or assume) that ρ_{uv} is close to unity and that $\sigma_u = \sigma_v$; i.e., Y is a very near simulacrum of X. In that case, we take

$$u^* = v. \tag{8}$$

This is perhaps the method of interpolation that is most widely used in actual statistical work, although any evidence that ρ_{uv} is close to unity and σ_u is close to σ_v is almost invariably absent. When these assumptions are not satisfied, this method may easily yield an interpolated value that is less rather than more reliable than Method L.[9]

(ii) *Method* R_2.[10] We make statistical estimates of ρ_{uv}, σ_u, σ_v for a single related series Y, compute from these a value of b as an estimate of β, and use equation 6 directly.

(iii) *Method* R_3.[11] This is a variant that arises when there are a number of related series that seem equally good, but not all of which are available for all the dates for which we propose to make estimates. This situation arose in our work in the form of a group of states (e.g., New England states) which, when tested, revealed reasonably homogeneous behavior. For each date and state in the group for which we need to interpolate we should like to use the information for as many other states in the group as possible.

Let $v_1, v_2, \ldots, v_n$ designate the relative deviations from trend for the n states for which information is known and u, as usual, the value it is desired to estimate. In general, the optimum way to estimate u

[8] Designated M_1 in Technical Paper 16.
[9] See Technical Paper 16, section II, pp. 9–14.
[10] Designated M_b in Technical Paper 16.
[11] This would also be designated M_b in the notation of Technical Paper 16, since it differs from R_2 only in the way b is estimated.

would be from a multiple regression of u on $v_1, \ldots, v_n$. But this requires much more information than is available, and the gap cannot be easily filled by estimation. A special case, in which the multiple regression reduces to a much simpler expression, is obtained by assuming that, while the v's may have different standard deviations (σ_{v_i} for the "true" but unknown value, s_{v_i} for an estimate), the correlation between any v_i and any other and also between any v_i and u is the same. Designate this common correlation coefficient by ρ for the "true" value, and r for an estimate of it.

Each v_i separately gives an estimate of u, and under our assumptions the estimates are equally good (because the variance of the estimate is $\sigma_u^2(1-\rho^2)$, and ρ is assumed the same for all the uv correlations). However, the mean or sum of these separate estimates of u will be a still better estimator. Let

$$v' = \frac{1}{n}\left(\frac{v_1}{\sigma_{v_1}} + \frac{v_2}{\sigma_{v_2}} + \cdots + \frac{v_n}{\sigma_{v_n}}\right)\sigma_u \tag{9}$$

be the estimator, constructed as the mean of the separate estimates, except that the common correlation coefficient is omitted.[12] The standard deviation of v' is then given by [13]

$$\sigma_{v'} = \frac{\sigma_u}{n}\sqrt{n[1+(n-1)\rho]}, \tag{10}$$

and the correlation between u and v' by [14]

$$\rho_{uv'} = \rho\,\frac{\sigma_u}{\sigma_{v'}} = \rho\left(\frac{n}{1+(n-1)\rho}\right)^{1/2}, \tag{11}$$

[12] For this mean to be the optimum estimator requires the further assumption that, aside from the common component used to estimate u, the different v's are independent.

[13] Proof: From (9)

$$\sigma_{v'}^2 = \frac{\sigma_u^2}{n^2}\left[\sum_{i=1}^{n}\frac{\sigma_{v_i}^2}{\sigma_{v_i}^2} + \sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\rho_{v_iv_j}\right]$$
$$= \frac{\sigma_u^2}{n^2}[n + n(n-1)\rho].$$

[14] Proof: Treat u's and v's as having zero means; then

$$\rho_{uv'} = \frac{Euv'}{\sigma_u\sigma_{v'}} = \frac{1}{\sigma_u\sigma_{v'}}\cdot\frac{\sigma_u}{n}\left[\frac{Euv_1}{\sigma_{v_1}} + \frac{Euv_2}{\sigma_{v_2}} + \cdots + \frac{Euv_n}{\sigma_{v_n}}\right]$$
$$= \frac{1}{n\sigma_{v'}}[\rho\sigma_u + \rho\sigma_u \ldots] = \frac{n\sigma_u}{n\sigma_{v'}}\cdot\rho = \frac{\sigma_u}{\sigma_{v'}}\cdot\rho.$$

hence the estimate of u from v' by Method R_3 is

$$u^* = \left(\rho \frac{\sigma_u}{\sigma_{v'}}\right) \frac{\sigma_u}{\sigma_{v'}} \cdot v', \tag{12}$$

or, substituting for $\sigma_{v'}$ its value from (10) and for ρ its estimate r,

$$u^* = \frac{nr}{1 + (n-1)r} \cdot v'. \tag{13}$$

In many applications of Method R_3 we have assumed that the standard deviations are the same for the different states, both the states entering into v' and the state to which the u refers. In that special case, equation 13 still holds but v' becomes simply the mean of the v's.

It may be worth noting that Method R_3, when it is applicable, is a very effective way of combining a number of pieces of information, each of which separately is fairly unreliable. For example, suppose ρ is 0.5, which means that a particular v_i would reduce the variance of the estimate of u by one-quarter. For $n = 6$, $\rho_{uv'}$ will be .72, which would reduce the variance of u by one-half.

Method of Estimating Relevant Parameters

Given a series Y to be used as the related series in Method R, we need to estimate the value of b to use. This problem is difficult because it cannot be estimated from the correlation between X and Y and their standard deviations at dates at which both X and Y are available but only from the correlation between u and v and their standard deviations, and both u and v are never available. We have therefore had to use a number of surrogates for u and v.

1. DIFFERENT TIME PERIODS. Sometimes both X and Y are known for the same time intervals for which interpolation is required but for a different period.[15] In that case, we have constructed u's and v's for the two series for the period for which both are available and used the correlation and regression coefficient between them as estimates for the period when X is not available.

[15] For example, before 1934, bank records for nonmember banks do not segregate U.S. government deposits from other deposits. However, beginning 1934, FDIC reports show this segregation for insured nonmember banks at semiannual dates. See Chapter 17 for a description of the way in which we used the relationships for the period after 1934 to estimate those for the period before.

2. DIFFERENT GEOGRAPHIC UNITS. The most extensive example of the use of different units is in connection with Method R_3. For example, we know mutual savings deposits in all relevant states for a June date each year but for only some states for inter-June dates. For those stretches of dates and states for which we have the requisite information we computed v's by expressing the inter-June values as relative deviations from trend values on a linear trend connecting two successive June values. We then correlated these v's for different states and computed their standard deviations. To estimate the common ρ that is required for Method R_3 we averaged the r's for the states that we regard as in the same group.[16]

3. DIFFERENT SERIES. Sometimes we have used the relation between two different series. For example, we used the relation between demand deposits less duplications (see section 7, below, for the definition) in *member* banks in one set of states and another set to estimate the value of b for interpolating demand deposits less duplications in *nonmember* banks in one of these sets of states from known values for nonmember banks in other states.

4. DIFFERENT TIME INTERVALS. This has been perhaps our most frequent recourse. For example, we have two series at June dates, and only one at inter-June dates. To estimate the b to use we have constructed u's and v's by treating Junes separated by two years as corresponding to the dates t_0 and t_2 and the intermediate June as corresponding to t_1. We have then estimated the correlations and standard deviations for these hypothetical u's and v's and regarded them as applying to the inter-June dates.

Choice of Related Series

Generally we have tested alternative series as possible interpolators. We have estimated r_{uv} for each by one of the methods just listed, the same method of course for all of the alternatives considered. We have selected the series which gave the highest r as the best of the alternatives.

[16] The average was constructed by first converting the r's to Fisher's Z [$Z = \frac{1}{2}\{\log_e(1 + r) - \log_e(1 - r)\}$, averaging the Z's, and then transforming back to r.

If that r was very high and the standard deviations of u and v seemed reasonably close, we have used Method R_1. If that r was very low and we were not in a position to use Method R_3, we have preferred to use Method L. We did not have a single numerical criterion for this choice, but it is probably fair to say that we tended to prefer Method L to Method R_2 if the indicated correlation was less than 0.5. One of the findings that surprised us most was how difficult it is to obtain a related series giving a correlation as high as 0.5. The contrary impression of statistical workers, reflected in their widespread use of Method R_1, derives, we believe, from their tendency to judge the correlation by looking at X and Y for dates when both are known rather than looking at counterparts to u and v.

Method S

Step Method Interpolation, which we shall designate *Method S,* is a device for interpolating by steps between values of X known at times t_0 and t_1 to obtain desired values of X at intervening time units. Under this method, if an even number of time units intervenes between t_0 and t_1, the value of X at t_0 is carried forward halfway and the value for X at t_1 is carried backward halfway. If the number of time units intervening between t_0 and t_1 is odd, all but the center unit are assigned the nearer of the two known values and the center one is made the average of the two known values.

An example of the one use we made of Method S in this volume is given in Chapter 11, where we applied it to convert a call date series into a monthly series. We proceeded as follows. The number of days intervening between a given call date and the following one was divided in half. This half-way date usually fell in the month following the given call date. The number of days in the month up to this half-way date was then expressed as a fraction of the total number of days in that month, the balance as 1.00 minus that fraction. For example, there are call dates on January 7 and April 1, 1867. The midpoint date is February 17.5, 1867. Since there are twenty-eight days in the month, 17.5 represents 0.625, the balance of 10.5 days, 0.375. The value for February was then taken to be 0.625 of the January call date figure plus 0.375 of the April call date figure. The months for which no figures were derived as just described were assigned the value of the nearest call date. In the above example, the value for March was the call date figure for April.

6. Reliability of the Estimates

Errors in our estimates arise from two major sources: errors in the original data underlying the estimates; and the interpolation and other devices that we have adopted to fill in gaps and correct other defects of coverage in the original data. One indication of the seriousness of the second source of error for our final call date and monthly estimates is the fraction of the final total derived from original data by interpolation according to Method L and according to Method R. Thus at the end of each of the following chapters, we give a summary table reporting the average fractions of each component of our estimates in these categories. These tables, a by-product of our estimates, provide a convenient summary of their structure, and the reader may find it helpful to consult them first, as a guide to the detailed explanation in the body of each chapter.

The fraction shown as interpolated is a substantial overestimate since it applies not to every date but only to those dates for which interpolation is required. For example, there are generally five or six call dates a year before 1923 (four, thereafter). Hence, in proceeding from call date to monthly estimates, additional information is required for roughly every other month before 1923 and for two months in each quarter thereafter. When this additional information is derived by interpolation, we have recorded the corresponding percentage of the estimate as interpolated in the reliability tables. The only offset to this overstatement is that our bench-mark data for June dates from 1896 on are treated as embodying no interpolation, whereas they may at times have involved some. But this is surely a minor offset.

As this remark indicates, the figures with the least error are almost surely the estimates for June each year. The maximum error is in the estimated change from one time unit to another between June dates. Our methods of interpolation have a bias in this respect. Undoubtedly they smooth the pattern of change during the year, so that the amplitude of actual inter-June movements is larger than of our estimated movements. This bias is in the absolute magnitude (absolute deviation or standard deviation) of the movements. Our methods have been designed to avoid as far as possible any bias in the algebraic average value of the inter-June movements (this is assured by our consistent use of deviations from June-to-June trends as the magnitude to be estimated). The bias in amplitude is a price that must be paid to minimize the fraction of the reported amplitude that is spurious; the bias is a result of our ignorance. For example, straight-line interpolation, on which we rely when our information on movements between dates for which we have evidence is inadequate, reduces the amplitude of reported figures around the trend between the dates for which we have estimates to zero, which is surely less than the actual amplitude.

It is difficult to say much that is illuminating about the first source of error—the errors in the original data underlying the estimates. Several things are clear, however. First, the final estimates are almost always the sum of a very large number of components, since the basic data are generally for individual banks, summed into totals for all national banks or all member banks, or all banks in a particular state and then summed over states. This means that there is much room for the law of large numbers to operate: the error in the total in relative terms must be very much smaller than any errors in the individual components that have a measure of statistical independence. Second, many of the errors in the individual components are independent; for example, simple reporting errors by individual banks, or classification errors by authorities in different states. Third, the reporting units that have compiled the data have done so for their own business purposes and hence have a strong interest in making them accurate. Indeed, the reason we are so skeptical of the breakdown of deposits between demand and time deposits before 1914 is precisely that this condition does not hold for these data prior to that year.

These three factors together lead us to believe that the aggregate figures have negligible residual errors arising simply from error in recording or reporting the figures summed. The significant errors in the aggregate reported figures must arise from errors that are not subject to cancellation: incompleteness of coverage of either reporting units or items reported, misclassifications for our purpose common to many units reporting (e.g., inclusion of items in process of collection with vault cash), errors in printing of final aggregates we have used, etc. Needless to say, an effort has been made to eliminate any known sources of error of this kind. But some unquestionably remain.

The nature of this source of error suggests that it affects the level of the estimates much more than the year-to-year, call-date-to-call-date, or month-to-month movements, since errors of this type are likely to have high serial correlation.

7. Definitions

These definitions are for our vault cash and currency estimates through 1942 and our deposit estimates through 1945, as originally constructed by us. In combining the estimates to obtain the consolidated monetary totals shown in Chapter 1, the definitions listed below were altered to make them applicable to the coverage of each consolidated total. For example, in Part Three we treat mutual savings bank vault cash as a component of bank vault cash and do not include it in currency held by the public. In Chapter 1, however, for the consolidated monetary totals restricted to commercial banks (Table 1, columns 8 and 9), mutual savings banks are treated as part of the public, and hence their vault cash is included in the public's currency holdings. For the consolidated monetary total covering commercial banks, mutual savings banks, and the Postal Savings System, vault cash holdings of all three types of institutions are excluded from the public's currency holdings (Table 1, column 11). The reader should also consult Chapter 1 for variations in the definitions listed below in years subsequent to 1945 covered by Federal Reserve estimates.

The public includes individuals, business firms other than banks, municipalities, states, and federal government agencies other than the Treasury Department in Washington, D.C., and the mints and assay offices in the country. For currency estimates the U.S. public includes the public and banks in foreign countries that own U.S. currency. For deposit estimates the U.S. public includes the public of foreign countries with deposits at banks located in the continental United States. Banks in the U.S. possessions and mutual savings banks in the United States are treated as banks, not as part of the public, for the currency estimates, but as part of the public for deposit estimates.

Government refers to the Treasury Department in Washington, D.C., and the mints and assay offices in the country as well as other departments, bureaus, and officials of the United States. It does not include government corporations and credit agencies.

All banks include all commercial and mutual savings banks.

Commercial banks include banks operating under commercial bank, trust company, or stock savings bank charters; cash depositories; private banks engaged in deposit banking; and Morris Plan and industrial banks operating under general banking codes. Beginning in 1920 foreign branches of commercial banks are excluded.

Mutual savings banks include all savings banks organized without stock, managed by a board of trustees, with earnings distributed among depositors as dividends (usually limited by law to a prescribed maximum), and surplus carried to the guaranty fund.

National banks are commercial banks incorporated under federal law operating under the supervision of the United States Comptroller of the Currency. All national banks in the continental United States are required by law to be members of the Federal Reserve System and of the Federal Deposit Insurance Corporation.

Nonnational banks are commercial banks chartered under state laws and subject to state supervision, and unincorporated (private) banks. Since the start of the Federal Reserve System, nonnational banks include member and nonmember commercial banks without national charters.

Member banks of the Federal Reserve System include national banks in the continental United States and state member banks.

State member banks are commercial banks chartered under state laws—together with a few mutual savings banks—that have been admitted to membership in the Federal Reserve System upon complying with certain prescribed conditions. All state member banks are required to be members of the FDIC and are subject to both federal and state supervision.

Nonmember banks are commercial banks that are not members of the Federal Reserve System. They include *insured nonmember banks,* which have been admitted to Federal deposit insurance upon meeting certain prescribed conditions, and *noninsured nonmember banks.*

General depositaries of government funds, formerly national banks only, now include insured domestic banks and insular territorial and foreign banks in which agents of the federal government deposit funds collected.

Special depositaries of government funds include qualified incorporated banks or trust companies and occasionally mutual savings banks or dealers in government securities, authorized by the Secretary of the Treasury to purchase government securities for their own or their customers' accounts and make payment by crediting the amount of the subscription to special government accounts.

Currency in circulation comprises currency held by the public and bank vault cash. It excludes currency held by the Treasury and, since 1914, by Federal Reserve Banks.

Currency held by the public comprises all kinds of publicly held bank, Federal Reserve Bank, or Treasury issues of coin or paper money, plus any such currency that has been carried abroad, lost, or destroyed.

Bank vault cash excludes the reporting national banks' *own* bank notes and includes all other kinds of U.S. currency, metallic and paper, issued by all banks, Federal Reserve Banks, and the Treasury—whether or not considered part of legal reserves—either held on the premises of member and nonmember banks or in transit to or from Federal Reserve Banks. It includes cash held by all commercial banks in the United States and its possessions and also in all mutual savings banks.

Government balances include government deposits at Federal Reserve Banks and their branches and at commercial and mutual savings banks and Treasury cash.

Treasury cash includes currency assets of the Treasury Department in Washington, D.C., and the mints and assay offices in the country, exclusive of the reserve held against gold and silver certificates, cash held for Federal Reserve Banks, and the gold redemption fund for Federal Reserve notes.

Total deposits adjusted is the sum of demand deposits adjusted and time deposits adjusted at commercial banks.

Demand deposits adjusted include all demand deposit items at commercial banks, except interbank demand deposits and U.S. government demand deposits, less cash items in process of collection.

Time deposits adjusted include all time deposit items at commercial banks except interbank time deposits, postal savings redeposited in banks, and U.S. government time deposits. (For a list of demand and time deposit items and cash items in process of collection, see *All-Bank Statistics, 1896–1955,* pp. 87–88.)

Demand deposits less duplications include all demand deposit items at commercial banks except interbank demand deposits and cash items in process of collection.