Have Postwar Economic Fluctuations Been Stabilized?

By Francis X. Diebold and Glenn D. Rudebusch*

Arthur F. Burns (1960 p. 2) was one of the first to assert that business cycles in the postwar era had changed in character:

> Between the end of the Second World War and the present, we have experienced four recessions, but each was a relatively mild setback. Since 1937 we have had five recessions, the longest of which lasted only thirteen months. There is no parallel for such a sequence of mild—or such a sequence of brief—contractions, at least during the past hundred years in our own country.

The steady growth of the 1960's produced a general acceptance of the view that the U.S. economy was more stable in the years after World War II than in the prewar period. This consensus was reinforced by formal examinations of postwar stabilization, notably by Martin N. Baily (1978) and J. Bradford De Long and Lawrence H. Summers (1986). Such examinations focused on the changing *volatility* of business fluctuations, and they uniformly concluded that the variability of various macroeconomic aggregates about trend had diminished during the postwar period.

The consensus on the postwar volatility stabilization of macroeconomic aggregates was seriously challenged by Christina D. Romer (1986a–c, 1988, 1989). She argued that the apparent higher volatility displayed by prewar aggregates (whether real gross national product [GNP], industrial production, or the unemployment rate) reflected differences in the methods used to construct prewar and postwar data; when similar methods are employed for both periods, she argued, the difference between prewar and postwar volatility is greatly lessened. In Romer's interpretation, the apparent postwar moderation of the business cycle was simply an artifact of inconsistent data.

Romer's contention has itself been challenged. Some authors have constructed still more alternative versions of prewar aggregates and have reached traditional conclusions about prewar versus postwar macroeconomic volatility (David R. Weir, 1986; Nathan S. Balke and Robert J. Gordon, 1989). Others, such as Stanley Lebergott (1986), have argued that Romer's reconstructed aggregates, like the original series, depend importantly on unverifiable assumptions and therefore are not unambiguously better than the original series. Our reading of the literature on volatility stabilization is that the paucity of source data makes it very difficult to construct incontrovertible aggregate measures of the prewar U.S. economy, even at the annual frequency. Moreover, because the quantitative size of fluctuations in these constructed macroeconomic aggregates will be crucial for the resolution of the volatility debate, the inadequacy of aggregate measures of the prewar economy undermines any comparison of prewar and postwar volatility.

Hence, we address the issue of stabilization, but we do not join the debate on volatility. Instead, we provide new evidence on the stability of the postwar economy by

*Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297 and Division of Monetary Affairs, Board of Governors of the Federal Reserve System, Washington, DC 20551, respectively. We thank Jeff Miron, Andy Rose, Mark Watson, David Wilcox, and two anonymous referees for comments. We gratefully acknowledge financial support from the National Science Foundation (grant SES 89-21715), the University of Pennsylvania Research Foundation (grant 3-71441), and the Institute for Empirical Macroeconomics. Joon-Haeng Lee, Tom Brennan, and Ralph Bradley provided research assistance. The views expressed here are our own and do not necessarily coincide with those of the Board of Governors of the Federal Reserve or its staff.

investigating a different aspect of stabilization and by employing a different type of data. Drawing upon the perspective of Diebold and Rudebusch (1990), we approach the question of stabilization in terms of the relative duration, rather than the relative volatility, of prewar and postwar business cycles. Duration is clearly one aspect of the postwar stabilization that Burns had in mind when he noted the unusual brevity, as well as mildness, of postwar contractions.[1] In modern terminology, the duration perspective considers the frequency of business cycles, while the volatility debate has focused only on their amplitude.

To examine durations, we employ a chronology of business-cycle turning points. By eschewing examination of the amplitude of business fluctuations, we avoid relying on estimates of the quantitative movements of a prewar macroeconomic aggregate, which are critical to conclusions about volatility. Compared with an aggregate measure of economic activity, a business-cycle chronology contains *less* information because the chronology is only qualitative, not quantitative, and *more* information because the chronology can incorporate a greater variety and number of sources of cyclical information. The former attribute is obvious: designating turning points largely requires only a qualitative sense of the direction of general business activity. Thus, for example, concluding that the second quarter of 1894 was a cyclical peak is much easier than determining that real GNP rose x percent in the second quarter and fell y percent in the third quarter of that year.

At the same time, because only qualitative information is required, a business-cycle chronology can be constructed from a greater number of indicators of business activity than just the components of an aggregate measure such as real GNP or industrial production. For example, the business-cycle chronology of the National Bureau of Economic Research (NBER), which we use below, incorporates a wide variety of sources of cyclical information, including the price movements of stocks and other assets, as well as descriptive accounts of economic activity from historical business annals. Sources such as these have necessarily been ignored in the volatility-stabilization debate, which has focused on aggregate measures; thus, our use of the NBER business-cycle chronology implicitly brings new information to the debate about the changing nature of business fluctuations.[2]

In our analysis, however, we do not accept the NBER chronology unquestioningly. One clear truth in U.S. economic history is that the quantity and quality of economic data have increased markedly over the last century. The relative scarcity and poor quality of earlier data may affect the comparability of prewar and postwar turning-point dates. Such data considerations may be important for judging changes in cyclical duration, just as similar data problems were crucial for the volatility debate. Accordingly, we take care to assess the robustness of our results to variations in the prewar chronology.

The paper proceeds as follows. In Section I, we discuss the NBER business-cycle dating procedures and the historical consistency of the NBER turning-point dates. In Section II, we describe a test of the null hypothesis of no duration stabilization, that

[1]Duration stabilization has largely been ignored by researchers; exceptions include De Long and Summers (1988), Victor Zarnowitz (1989), and Daniel E. Sichel (1991), who address similar issues with other techniques.

[2]The ability of the NBER to construct its chronology at a monthly frequency demonstrates the richness of the chronology's information set. Previous volatility studies have been able to construct the requisite aggregates at only an annual frequency, which is quite crude for assessing business cycles. Consideration of broader information sets in volatility comparisons also motivates the analyses of Matthew D. Shapiro (1988), who uses stock prices, and Steven M. Sheffrin (1988), who uses international data.

TABLE 1—NBER BUSINESS-CYCLE DATES AND DURATIONS

Trough	Peak	Contractions	Expansions	Trough to trough	Peak to peak
December 1854	June 1857	—	30	—	—
December 1858	October 1860	18	22	48	40
June 1861	April 1865	8	46	30	54
December 1867	June 1869	32	18	78	50
December 1870	October 1873	18	34	36	52
March 1879	March 1882	65	36	99	101
May 1885	March 1887	38	22	74	60
April 1888	July 1890	13	27	35	40
May 1891	January 1893	10	20	37	30
June 1894	December 1895	17	18	37	35
June 1897	June 1899	18	24	36	42
December 1900	September 1902	18	21	42	39
August 1904	May 1907	23	33	44	56
June 1908	January 1910	13	19	46	32
January 1912	January 1913	24	12	43	36
December 1914	August 1918	23	44	35	67
March 1919	January 1920	7	10	51	17
July 1921	May 1923	18	22	28	40
July 1924	October 1926	14	27	36	41
November 1927	August 1929	13	21	40	34
March 1933	May 1937	43	50	64	93
June 1938	February 1945	13	80	63	93
October 1945	November 1948	8	37	88	45
October 1949	July 1953	11	45	48	56
May 1954	August 1957	10	39	55	49
April 1958	April 1960	8	24	47	32
February 1961	December 1969	10	106	34	116
November 1970	November 1973	11	36	117	47
March 1975	January 1980	16	58	52	74
July 1980	July 1981	6	12	64	18
November 1982	July 1990	16	92	28	108

Note: Durations are given in months. Wartime expansions and whole cycles are underlined.

is, that the distributions of prewar and postwar durations are identical. We provide empirical results in Section III and offer a summary and interpretation in Section IV.

I. The NBER Business-Cycle Chronology

The dates of U.S. business-cycle peaks and troughs designated by the NBER are shown in Table 1, along with the associated durations of expansions, contractions, and whole cycles (measured from peak to peak and from trough to trough). As noted above, the earlier volatility debate has hinged on the issue of the comparability of prewar and postwar data, and we focus the discussion in this section on an analogous issue: the historical consistency of the prewar and postwar NBER turning-point dates and the comparability of the associated cyclical durations.

A brief review of the NBER dating procedure is in order. An early description of this method is Burns and Wesley C. Mitchell (1946 pp. 76–7):[3]

> Our first step toward identifying business cycles was to identify the turns of general business activity indicated by [descriptive business] annals. Next, the

[3] A more recent description is Geoffrey H. Moore and Zarnowitz (1986), which provides an excellent overview of the NBER cyclical dating method and related issues.

> evidence of the annals was checked against indexes of business conditions and other series of broad coverage. In most cases these varied records pointed clearly to some one year as the time when a cyclical turn occurred. When there was conflict of evidence, additional statistical series were examined and historical accounts of business conditions consulted, until we felt it safe to write down an interval within which a cyclical turn in general business probably occurred. We then proceeded to refine the approximate dates by arraying the cyclical turns in the more important monthly or quarterly series we had for the time and country.

The last step is the most important, because it focuses directly on the amount of cyclical comovement or coherence among economic variables. For Burns and Mitchell, this comovement is the prime definitional characteristic of the business cycle: "...a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions..." (Burns and Mitchell, 1946 p. 3). Thus, in determining the monthly dates of business-cycle turning points, Burns and Mitchell considered hundreds of individual series, including those measuring commodity output, income, prices, interest rates, banking transactions, and transportation services. The turning points of these individual series are not randomly distributed; rather, they form clusters of peaks and troughs. The monthly dates of the central tendencies of such clusters are designated as the turning points of the general business cycle. For the period from 1854 through 1938, these dates are listed by Burns and Mitchell (1946 p. 105). Dates in the postwar period have been designated by successive NBER researchers who have closely adhered to the Burns and Mitchell methodology (see Moore and Zarnowitz, 1986).[4] Note that, contrary to popular folklore, NBER researchers have never used two consecutive quarterly declines in real GNP as the criterion for dating downturns.

The historical consistency of the procedures used by NBER researchers to designate turning points supports the use of these dates in prewar–postwar comparisons. Nevertheless, although the general dating procedures have not changed, both the number and quality of the underlying individual series examined have greatly increased over time. For example, in Burns and Mitchell's (1946 p. 82) analysis only 19 individual monthly or quarterly series were available for dating in the 1860's, while 199 were available for the dates after 1890, and 665 were available after 1920. The increase in the number of underlying individual series, which was also accompanied by an increase in the quality of most series, is presumably associated with increased reliability of the NBER dates. Clusters of individual turning points are quite narrow in the postwar period; in contrast, inadequate data result in much more uncertainty about some of the prewar NBER dates. The changes in the reliability of the dates, as certain individual series necessarily assume more importance in the absence of others in the prewar period, could affect the validity of a prewar–postwar comparison of NBER cyclical durations. The rest of this section addresses this issue and describes some of the variations of the canonical NBER chronology that we consider in order to ensure the robustness of our results.

All of the researchers who have designated NBER turning points have cautioned that there is some uncertainty about the precise timing of the general turns in business activity. One indication of the uncertainty associated with the official dates is the discrepancy between these dates and a number of alternative dates that have been suggested by NBER researchers and by independent observers.[5] Let us first consider

[4]Two detailed illustrations of the postwar application of the NBER dating methodology are Zarnowitz and Moore (1977) for the 1973–1975 recession and Zarnowitz and Moore (1983) for the 1980 recession.

[5]Indeed, this is one of the procedures used by Burns and Mitchell (1946 p. 108) to examine the dependability of their dates.

the reliability of the postwar dates. The NBER turning-point dates during the early part of the postwar period were the subject of some controversy, with several alternative chronologies hotly debated (Moore, 1961; Lorman C. Trueblood, 1961; George W. Cloos, 1963a, b; Zarnowitz, 1963a, b). The differences between the proposed alternatives and the official postwar chronology are minor; of the eight dates examined by Cloos, for example, his suggested changes would shift one peak back by one month, another forward by two months, and one trough back by three months. Given the striking nature of our subsequent results, these differences are insignificant.

The choice of more recent dates in the postwar period (since 1960), and indeed the entire NBER turning-point methodology, has gained additional support from research by James H. Stock and Mark W. Watson (1989).[6] They have attempted to formalize the notion that the business cycle is defined by the comovements of many macroeconomic time series by specifying a dynamic factor model that identifies the unobserved common component in the movements of many coincident variables. The cyclical peaks and troughs of the extracted common component coincide with the NBER chronology, except in 1969, when the NBER-dated peak is two months later.

As suggested above by the large changes in the number of time series employed by Burns and Mitchell (1946), the prewar dates are of varying quality. The dates in the interwar period (1918–1938) appear to be little more questionable than those in the postwar period. Of the original 12 turning points in this period specified by Burns and Mitchell (1946), careful reevaluations by the NBER staff led to three changes of one month and two shifts of two months (Moore and Zarnowitz, 1986). These revisions are broadly indicative of the small amount of uncertainty in the interwar dates.

[6]The postwar NBER chronology is also broadly confirmed by James Hamilton (1989), who posits an underlying nonlinear regime-switching model and uses optimal-signal-extraction techniques to estimate turning-point dates.

The turning-point dates before World War I are more questionable. Again, we can compare alternative business-cycle chronologies for this period, such as those of Joseph Kitchin (1923), Warren M. Persons (1931), and Leonard Ayres (1939), in order to gauge the uncertainty associated with the NBER's choices. From this perspective, the NBER dates appear to be reasonable choices, with no clear bias; however, the range in variation among the alternatives is fairly large, with an average shift of about four months. Careful examinations of the early NBER dates, notably Rendigs Fels (1959) and Zarnowitz (1981), place the greatest uncertainty on the timing of the dates before 1885. Very few comprehensive statistics are available at a monthly frequency before the mid-1880's; consequently, the clusters of individual series available for Burns and Mitchell (1946) are rather sparse and diffuse. In our empirical analysis, we shall examine the robustness of our results when the pre-1885 turning points are excluded.

Although the early NBER dates appear to provide a reasonably unbiased delineation of good times from bad, there is a remaining question about whether some of the designated recessions represent true cyclical contractions or rather are simply periods of very slow growth (i.e., growth recessions). This distinction is more difficult to make for recessions in the pre-World War I period because several data series are only available on a trend-adjusted basis, making actual declines in real economic activity difficult to judge. In the period after 1885, the 1887–1888 recession is the most dubious, although the 1899–1900 recession was also very mild (Kitchin, 1923; A. Ross Eckler, 1933; Fels, 1959; Zarnowitz, 1981). Although we remain undecided on the classification of these episodes, we examine the consequences of treating 1887–1888 and 1899–1900 as growth slowdowns rather than as business-cycle contractions.

In light of the above concerns about the historical consistency of the NBER dates, we consider two variations on the official chronology in order to assess the robustness of our results: (i) exclusion of the pre-1885

TABLE 2—LISTING OF DURATION SAMPLES

A. *Pre-World War II (December 1854–June 1938):*	
A1:	All observations
A2:	Excluding observations before May 1885
A3:	A2, eliminating 1887 and 1899 contractions
A1*:	A1, excluding wartime observations
A2*:	A2, excluding wartime observations
A3*:	A3, excluding wartime observations
B. *Pre-Great Depression (December 1854–August 1929):*	
B1:	All observations
B2:	Excluding observations before May 1885
B3:	B2, eliminating 1887 and 1899 contractions
B1*:	B1, excluding wartime observations
B2*:	B2, excluding wartime observations
B3*:	B3, excluding wartime observations
C. *Pre-World War I (December 1854–December 1914):*	
C1:	All observations
C2:	Excluding observations before May 1885
C3:	C2, eliminating 1887 and 1899 contractions
C1*:	C1, excluding wartime observations
C2*:	C2, excluding wartime observations
C3*:	C3, excluding wartime observations
D. *Post-World War II (February 1945–July 1990):*	
Z:	All observations
Z*:	Z, excluding wartime observations

turning-point dates in order to avoid potentially unreliable dates in the very early period and (ii) elimination of the 1887 and 1899 recessions[7] in order to account for the possibility that these were merely growth recessions. As a further sensitivity test, we consider three different terminal dates for the prewar period (June 1938, August 1929, and December 1914), thus excluding from consideration the Great Depression and other interwar recessions, which may be atypical observations. Finally, we also consider the exclusion of wartime expansions and cycles in order to avoid possible spuriously long observations. A complete listing of all of the various duration samples used in our analysis is given in Table 2, along with the associated mnemonics. (The A, B, and C samples are all loosely termed "prewar" samples.)

II. A Test of Duration Stabilization

Consider the two samples of prewar and postwar durations of size n_x and n_y, $\{X_1,\dots,X_{n_x}\}$ and $\{Y_1,\dots,Y_{n_y}\}$. Denote the corresponding population prewar and postwar duration distribution functions by F and G. The null hypothesis of no postwar duration stabilization implies that these distributions are identical ($F = G$). Depending on the situation, we shall subsequently be interested in both one-sided and two-sided alternatives. The interpretation of the one-sided alternative that Y is stochastically larger than X is that (i) $F \neq G$ and (ii) $G(k) \leq F(k)$ for all k [or equivalently, $P(Y > k) \geq P(X > k)$ for all k]. The inequalities are reversed for the one-sided alternative that X is stochastically larger than Y. The two-sided alternative, $F \neq G$, has the obvious interpretation.

We shall test the null hypothesis of no postwar stabilization using the Wilcoxon, or rank-sum, test. Replace the observations $\{X_1,\dots,X_{n_x},Y_1,\dots,Y_{n_y}\}$ by their ranks, $\{R_1,\dots,R_n\}$, where $n = n_x + n_y$.[8] Then the Wilcoxon test statistic is formed as the sum of the ranks in the second sample:

$$(1) \qquad W = \sum_{i=n_x+1}^{n} R_i.$$

The intuition of this statistic is obvious: under the null hypothesis that $F = G$, the average rank of an observation in the prewar sample should equal the average rank of an observation in the postwar sample, and W is a sufficient statistic for this comparison. Furthermore, the distribution of W under the null hypothesis that $F = G$ is invariant to the underlying distribution of durations. This invariance follows from the fact that the null distribution of the ranks (assuming the independence of observations) is simply given by

$$(2) \qquad P(R_1 = r_1,\dots,R_n = r_n) = 1/n!$$

[7] The "elimination" of a recession means that we replace that contraction and its immediate preceding and succeeding expansions by one long expansion.

[8] In the case of a tie, the relevant ranks are replaced by the average of the ranks of the tied observations.

for all permutations $(r_1,\dots,r_n)$ of $(1,\dots,n)$. Because W is a function of the ranks, the distribution of W is also invariant to the underlying distribution of durations. Indeed, equation (2) enables computation of exact finite-sample p values of W, which are calculated numerically using the algorithm of Diebold et al. (1992).[9]

The Wilcoxon test is a nonparametric test designed to have particularly high power against alternatives involving a shift of location. Intuition on this point can be gained by comparing the Wilcoxon test statistic to the classical t statistic for testing equality of two population means,

$$(3)\qquad t=(n_x n_y/n)^{1/2}\left[(\bar{Y}-\bar{X})/s\right]$$

where

$$(4)\qquad s=(n-2)^{-1/2}\left[\sum_{i=1}^{n_x}(X_i-\bar{X})^2+\sum_{j=1}^{n_y}(Y_j-\bar{Y})^2\right]^{1/2}$$

The t statistic is appropriate for testing the null hypothesis that $E(X)=E(Y)$ when the underlying populations are normally distributed. Unfortunately, normality is a distinctly inappropriate distributional assumption for duration data. The Wilcoxon test may be interpreted as a distribution-free t test, obtained by replacing the observations $\{X_1,\cdots,X_{n_x},Y_1,\cdots,Y_{n_y}\}$ by their ranks $\{R_1,\cdots,R_n\}$, which yields

$$(5)\qquad t^*=(n_x n_y/n)^{1/2}\left[\left(\bar{R}_y-\bar{R}_x\right)\Big/s^*\right]$$

where $\bar{R}_x$ and $\bar{R}_y$ denote the mean ranks of the X and Y samples, and

$$(6)\qquad s^*=(n-2)^{-1/2}\times\left[\sum_{i=1}^{n_x}(R_i-\bar{R}_x)^2+\sum_{j=n_x+1}^{n}\left(R_j-\bar{R}_y\right)^2\right]^{1/2}.$$

Straightforward but tedious algebra reveals t^* to be a monotonic transformation of W.

Because the Wilcoxon test is exact, we are assured of correct test size, even in small samples. Surprisingly, the test also has good power against a variety of alternatives. The trade-off between the relaxation of distributional assumptions and the loss of power is extremely favorable: the Wilcoxon test is only slightly less powerful than the t test when the distributional assumption (normality) underlying the t test is true, and it may be much more powerful when the distributional assumption is false.[10]

Under the maintained assumption that the distributions of durations differ only by a shift in location [i.e., $G(k)=F(k+\Delta)$ for all k], we can also produce a confidence interval for the location shift, Δ. Consider the $n_x n_y$-element sequence of differences $\{D_{ij}\}$, $i=1,\dots,n_x$, $j=1,\dots,n_y$, where $D_{ij}=Y_j-X_i$, and order them so that $D_{(1)}<D_{(2)}<\cdots<D_{(n_x n_y)}$. For a given significance level α, let k_α be an integer defined from the confidence interval

$$(7)\qquad P(k_\alpha\le U\le n_x n_y-k_\alpha)=1-\alpha$$

where

$$(8)\qquad U=W-n_y(n_y+1)/2$$

[9]Critical values are also tabulated in John V. Bradley (1968) for $n_x,n_y\le 25$.

[10]See Peter J. Bickel and Kjell A. Doksum (1977) for a discussion of the comparative performance of the Wilcoxon and t tests.

is the Mann-Whitney U statistic, a monotonic transformation of W.[11] Then it can be shown (Bickel and Doksum, 1977) that

$$(9)\quad P\left(D_{(k_\alpha)}\right) \leq \Delta \leq D_{(n_x n_y - k_\alpha + 1)} = 1 - \alpha.$$

Thus, a two-sided $(1-\alpha)$-percent confidence interval for Δ is $(D_{(k_\alpha)}, D_{(n_x n_y - k_\alpha + 1)})$. Alternatively, the two $(1-\alpha)$-percent one-sided confidence intervals are $(D_{(k_{2\alpha})}, \infty)$ and $(-\infty, D_{(n_x n_y - k_{2\alpha} + 1)})$.

III. Empirical Results

Before applying the Wilcoxon test, we first must verify two features of the data in order to ensure the validity of the testing procedure: first, the independence of duration observations and, second, the constancy of trend growth in the prewar and postwar periods. The independence assumption, which was required to obtain appropriate critical values for the Wilcoxon test, appears to be a good working assumption. The correlations between the lengths of successive expansions or between the lengths of successive contractions (over the entire sample) are insignificantly different from zero at even the 20-percent level.

The second pretest issue reflects the fact that business cycles are delineated on a non-trend-adjusted basis; thus, any differences in the trend growth of the economy in the prewar and postwar periods would affect duration comparisons. If the postwar economy had a higher average rate of growth than the prewar economy and each economy had identical trend-adjusted cyclical movements, the duration of postwar expansions would be longer and the duration of postwar contractions would be shorter than their prewar counterparts. However, as shown in Table 3, the mean growth rate of real output in the postwar period was little different than in the prewar period. (The prewar growth rates are calculated over several ranges that roughly correspond to our prewar duration samples, whose mnemonics are given in parentheses in Table 3).[12] Thus, any evidence for duration stabilization does not reflect changes in trend growth.

Table 3—Mean Growth Rate of Real GNP

Sample	Mean ($\Delta \log Y_t$)
Postwar sample:	
1946–1989 (Z)	0.025
Prewar samples:	
1870–1938 (A1)	0.031
1886–1938 (A2)	0.027
1870–1929 (B1)	0.037
1886–1929 (B2)	0.034
1870–1914 (C1)	0.038
1886–1914 (C2)	0.033

Note: The real GNP sample from 1869 to 1929 comes from Romer (1989 pp. 22–3); the later data come from the national income and product accounts (NIPA).

With these two issues settled, results from the Wilcoxon tests for expansions and contractions appear in Tables 4 and 5, respectively. For each pair of prewar and postwar samples, we report sample sizes, mean durations, the Wilcoxon statistic and its one-sided p value, and approximate 90-percent and 80-percent one-sided confidence intervals for the location shift.[13] For example, the top row of Table 4 compares the prewar expansion sample A1 (with 21 observations and a mean duration of 26.5 months) and the postwar expansion sample Z (with nine observations and a mean duration of 49.9 months). For these two samples, the exact Wilcoxon p value under the null hypothesis of no change in distribution is less than 0.01, and the confidence-interval estimates suggest that we can be 90-percent certain

[11] The finite-sample distribution of U is tabulated in Bickel and Doksum (1977).

[12] Note that we rely on the prewar measure of GNP only for average growth estimates, rather than using it for the more contentious assessment of properties of cyclical fluctuations.

[13] The obvious alternatives of longer postwar expansions and shorter postwar contractions make one-sided tests and confidence intervals appropriate.

TABLE 4—WILCOXON TEST FOR EXPANSIONS

Sample		Sample size		Mean duration		Wilcoxon test		Confidence interval	
x	y	n_x	n_y	$\bar{x}$	$\bar{y}$	W	$P_1(W)$	90-percent	80-percent
A1	Z	21	9	26.5	49.9	193.5	0.006	< -9	< -12
A2	Z	15	9	24.7	49.9	154.0	0.006	< -12	< -14
A3	Z	13	9	30.8	49.9	127.5	0.055	< -3	< -8
A1*	Z*	19	7	24.5	42.6	132.5	0.013	< -5	< -9
A2*	Z*	14	7	23.3	42.6	106.0	0.015	< -6	< -12
A3*	Z*	12	7	29.8	42.6	85.5	0.098	< 0	< -4
B1	Z	20	9	25.3	49.9	190.5	0.004	< -10	< -14
B2	Z	14	9	22.9	49.9	151.0	0.003	< -13	< -15
B3	Z	12	9	29.3	49.9	124.5	0.035	< -4	< -12
B1*	Z*	18	7	23.1	42.6	130.5	0.007	< -6	< -10
B2*	Z*	13	7	21.2	42.6	104.0	0.007	< -10	< -12
B3*	Z*	11	7	27.9	42.6	83.5	0.063	< -3	< -6
C1	Z	15	9	25.5	49.9	154.5	0.005	< -9	< -12
C2	Z	9	9	21.8	49.9	56.0	0.004	< -12	< -15
C3	Z	7	9	32.4	49.9	47.5	0.105	< 4	< -4
C1*	Z*	14	7	24.0	42.6	106.5	0.012	< -5	< -7
C2*	Z*	9	7	21.8	42.6	80.0	0.016	< -6	< -12
C3*	Z	7	7	32.4	42.6	45.5	0.191	< 6	< 0

Note: Samples are identified in Table 2. The mean durations and the Wilcoxon test statistic are given in months. $P_1(W)$ is a one-sided p value for the null hypothesis of no postwar duration stabilization.

TABLE 5—WILCOXON TEST FOR CONTRACTIONS

Sample		Sample size		Mean duration		Wilcoxon test		Confidence interval	
x	y	n_x	n_y	$\bar{x}$	$\bar{y}$	W	$P_1(W)$	90-percent	80-percent
A1	Z	21	9	21.2	10.7	75.0	0.001	> 3	> 5
A2	Z	15	9	17.8	10.7	68.0	0.003	> 3	> 3
A3	Z	13	9	18.2	10.7	66.0	0.006	> 3	> 3
B1	Z	19	9	20.5	10.7	73.0	0.002	> 3	> 5
B2	Z	13	9	16.2	10.7	66.0	0.006	> 2	> 3
B3	Z	11	9	16.4	10.7	64.0	0.010	> 2	> 3
C1	Z	15	9	22.5	10.7	61.0	0.001	> 6	> 7
C2	Z	9	9	17.7	10.7	117.0	0.002	> 4	> 5
C3	Z	7	9	18.3	10.7	84.0	0.004	> 5	> 7

Note: Samples are identified in Table 2. The mean duration and the Wilcoxon test statistic are given in months. P_1W is a one-sided p value for the null hypothesis of no postwar duration stabilization.

that the postwar increase in mean expansion duration was at least 9 months. Results are shown for the other pairs of expansion samples in Table 4 and for contraction samples in Table 5. Almost without exception, the tests reject the null hypothesis of no stabilization in favor of longer postwar expansions or shorter postwar contractions. For contractions, rejection is always at the 1-percent level or better. For expansions, the evidence is slightly less overwhelming: 12 of 18 Wilcoxon p values for expansions are less than or equal to 0.02, but one sample rejects at only the 20-percent level, and two other samples reject at about the 10-percent level.

TABLE 6—WILCOXON TEST FOR PEAK-TO-PEAK CYCLES

Sample		Sample size		Mean duration		Wilcoxon test		Confidence interval	
x	y	n_x	n_y	$\bar{x}$	$\bar{y}$	W	$P_2(W)$	90-percent	80-percent
A1	Z	20	9	47.9	60.6	158.0	0.294	(−24,7)	(−19,3)
A2	Z	14	9	43.0	60.6	134.0	0.110	(−34,0)	(−26,−5)
A3	Z	12	9	46.8	60.6	115.0	0.278	(−33,9)	(−26,3)
A1*	Z*	18	7	46.6	53.3	101.5	0.534	(−18,10)	(−15,7)
A2*	Z*	13	7	41.2	53.3	89.5	0.210	(−32,7)	(−17,0)
A3*	Z*	11	7	45.0	53.3	75.5	0.426	(−28,11)	(−17,7)
B1	Z	19	9	45.6	60.6	156.0	0.224	(−28,4)	(−20,0)
B2	Z	13	9	39.2	60.6	132.0	0.060	(−38,3)	(−32,−6)
B3	Z	11	9	42.6	60.6	113.0	0.176	(−38,4)	(−30,−1)
B1*	Z*	17	7	43.8	53.3	100.5	0.418	(−22,8)	(−15,5)
B2*	Z*	12	7	36.8	53.3	88.5	0.120	(−33,2)	(−28,−3)
B3*	Z*	10	7	40.2	53.3	74.5	0.270	(−32,7)	(−27,3)
C1	Z	14	9	47.6	60.6	123.0	0.368	(−24,7)	(−19,4)
C2	Z	8	9	38.8	60.6	54.0	0.092	(−39,0)	(−32,−5)
C3	Z	6	9	45.0	60.6	40.0	0.388	(−38,11)	(−26,7)
C1*	Z*	13	7	47.2	53.3	78.5	0.700	(−18,11)	(−14,8)
C2*	Z*	8	7	38.6	53.3	66.5	0.232	(−34,7)	(−18,0)
C3*	Z*	6	7	45.0	53.3	38.5	0.628	(−19,14)	(−17,11)

Note: Samples are identified in Table 2. The mean durations and the Wilcoxon test statistic are given in months. $P_2(W)$ is a one-sided p value for the null hypothesis of no postwar change in the duration distribution.

Even more persuasive evidence is provided by a test of the joint hypothesis of both longer expansions and shorter contractions. Given a postwar duration stabilization that results in either (or both) longer expansions and shorter contractions, expansion-to-contraction ratios will be larger in the postwar period. In light of the separate results for expansions and contractions, it is not surprising that the Wilcoxon statistics for their ratios, which test a joint stabilization hypothesis, are generally less than 0.001. We interpret these results as the most compelling evidence supporting overall postwar duration stabilization.

It is unusual in empirical macroeconomics to obtain such strong results, particularly with small samples. But what of the more important question: are the postwar shifts significant from an *economic* perspective? Clearly, the answer is yes. Our results indicate that while less than 20 percent of the postwar period was spent in recession, more than 40 percent of the prewar period was spent in recession. Furthermore, the mean postwar expansion duration is *double* that of its prewar counterpart, while the mean postwar contraction duration is *half* that of its prewar counterpart.

The results are very different for whole cycles, whether measured from trough to trough or from peak to peak. Table 6 provides the statistics for cycles measured from peak to peak; similar results were obtained for trough-to-trough cycles. The p values of the Wilcoxon tests rarely indicate significant change in the postwar period; in fact, they are typically greater than 0.2. Thus, the data suggest an unchanged distribution of whole-cycle durations but with a revised allocation of time so that postwar expansions are longer, and contractions shorter.

IV. Summary

We have investigated the postwar-stabilization hypothesis from the perspective of *duration*, or frequency, as opposed to volatility, or amplitude. Our analysis made use of the qualitative information contained in the NBER's business-cycle chronology and was robust to criticisms of conventional measures of prewar aggregate data. Using a distribution-free statistical procedure, we found strong evidence of a postwar shift toward longer expansions and shorter contractions, which is consistent with a broad interpretation of the stabilization hypothesis. Moreover, we found no evidence for a postwar shift in the distribution of whole-cycle durations.

To the extent that postwar volatility was stabilized, one expects, ceteris paribus, concomitant duration stabilization due to the upward trend in aggregate economic activity. To see this, consider an extreme case: in an upwardly trending economy, as volatility approaches zero, expected expansion duration grows without bound, and expected contraction duration collapses to zero.[14] However, we believe that it is highly unlikely that all of the postwar duration stabilization is associated with volatility stabilization. To the extent that volatility actually was reduced, previous research has found that the reduction was small and hard to detect. The postwar shift toward duration stabilization, however, is large and difficult to deny. It is likely, therefore, that duration stabilization arose, at least in part, independently of volatility stabilization. Furthermore, some of the structural changes in the economy that have been cited as possible sources for volatility stabilization may actually impede duration stabilization. For example, it is fairly well established that the existence of a countercyclical entitlement program such as unemployment insurance *increases* individual unemployment durations by reducing the adverse effect of unemployment on personal income (e.g., Bruce D. Meyer, 1990). Such a program, although an "automatic stabilizer" in the sense of reducing the severity of contractions and the variability of fluctuations, may not generally shorten the durations of contractions or lead to duration stabilization.

[14]However, it should be stressed that the link between volatility stabilization and duration stabilization may be affected by other changes in the nature of business cycles, notably in the asymmetry of the cycle.

REFERENCES

Ayres, Leonard, *Turning Points in Business Cycles*, New York: Macmillan, 1939.

Baily, Martin Neal, "Stabilization Policy and Private Economic Behavior," *Brookings Papers on Economic Activity*, 1978, (1), 11–60.

Balke, Nathan S. and Gordon, Robert J., "The Estimation of Prewar Gross National Product: Methodology and New Evidence," *Journal of Political Economy*, February 1989, *97*, 38–92.

Bickel, Peter J. and Doksum, Kjell A., *Mathematical Statistics*, San Francisco: Holden-Day, 1977.

Bradley, John V., *Distribution-Free Statistical Tests*, Englewood Cliffs, NJ: Prentice-Hall, 1968.

Burns, Arthur F., "Progress Toward Economic Stability," *American Economic Review*, March 1960, *50*, 1–19.

______ **and Mitchell, Wesley C.,** *Measuring Business Cycles*, New York: National Bureau of Economic Research, 1946.

Cloos, George W., (1963a) "How Good are the National Bureau's Reference Dates?" *Journal of Business*, January 1963, *36*, 14–32.

______, (1963b) "More on Reference Dates and Leading Indicators," *Journal of Business*, July 1963, *36*, 352–64.

De Long, J. Bradford and Summers, Lawrence H., "The Changing Cyclical Variability of Economic Activity in the United States," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, 1986, pp. 679–719.

______ **and Summers, Lawrence H.,** "How Does Macroeconomic Policy Affect Output?" *Brookings Papers on Economic Activity*, 1988, (2), 433–80.

Diebold, Francis X. and Rudebusch, Glenn D., "A Nonparametric Investigation of Duration Dependence in the American Business Cycle," *Journal of Political Economy*, June 1990, *98*, 596–616.

______, ______ **and Tanizaki, Hisashi,** "On the Comparative Size and Power Properties of the Fisher, Wilcoxon, and *t* Tests," unpublished manuscript, University of Pennsylvania, 1992.

Eckler, A. Ross, "A Measure of the Severity of Depressions, 1873–1932," *Review of Economics and Statistics*, May 1933, *15*, 75–81.

Fels, Rendigs, *American Business Cycles, 1865–1897*, Chapel Hill: University of North Carolina Press, 1959.

Hamilton, James, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, March 1989, *57*, 357–84.

Kitchin, Joseph, "Cycles and Trends in Economic Factors," *Review of Economics and Statistics*, January 1923, *5*, 10–16.

Lebergott, Stanley, "Discussion" [Diffusion and Timing of Business Cycle Turning Points], *Journal of Economic History*, June 1986, *46*, 367–71.

Meyer, Bruce D., "Unemployment Insurance and Unemployment Spells," *Econometrica*, July 1990, *58*, 757–82.

Moore, Geoffrey H., "Discussion," *Proceedings of the ASA, Business and Economic Statistics Section, 1961*, Washington, DC: American Statistical Association, 1961, pp. 34–7.

______ **and Zarnowitz, Victor,** "The Development and Role of the NBER's Business Cycle Chronologies," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, 1986, pp. 735–79.

Persons, Warren M., *Forecasting Business Cycles*, New York: Wiley, 1931.

Romer, Christina D., (1986a) "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, February 1986, *94*, 1–37.

______, (1986b) "Is the Stabilization of the Postwar Economy a Figment of the Data?" *American Economic Review*, June 1986, *76*, 314–34.

______, (1986c) "New Estimates of Prewar Gross National Product and Unemployment," *Journal of Economic History*, June 1986, *46*, 341–52.

______, "World War I and the Postwar Depression: A Reinterpretation Based on Alternative Estimates of GNP," *Journal of Monetary Economics*, July 1988, *22*, 91–115.

______, "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869–1908," *Journal of Political Economy*, February 1989, *97*, 1–37.

Shapiro, Matthew D., "The Stabilization of the U.S. Economy: Evidence from the Stock Market," *American Economic Review*, December 1988, *78*, 1067–79.

Sheffrin, Steven M., "Have Economic Fluctuations Been Dampened? A Look at Evidence Outside the United States," *Journal of Monetary Economics*, January 1988, *21*, 73–83.

Sichel, Daniel E., "Business Cycle Duration Dependence: A Parametric Approach," *Review of Economics and Statistics*, May 1991, *73*, 254–60.

Stock, James H. and Watson, Mark W., "New Indexes of Coincident and Leading Economic Indicators," in O. Blanchard and S. Fischer, eds., *NBER Macroeconomics Annual, 1989*, Cambridge, MA: MIT Press, 1989, pp. 351–94.

Trueblood, Lorman C., "The Dating of Postwar Business Cycles," *Proceedings of The ASA, Business and Economic Statistics Section, 1961*, Washington, DC: American Statistical Association, 1961, pp. 16–26.

Weir, David R., "The Reliability of Historical Macroeconomic Data for Comparing Cyclical Stability," *Journal of Economic History*, June 1986, *46*, 353–65.

Zarnowitz, Victor, (1963a) "On the Dating of Business Cycles," *Journal of Business*, April 1963, *36*, 179–99.

______, (1963b) "Cloos on Reference Dates and Leading Indicators: A Comment," *Journal of Business*, October 1963, *36*, 461–3.

______, "Business Cycles and Growth: Some Reflections and Measures," in W. J. Muckl and A. E. Ott, eds., *Wirtschaftstheorie und Wirtschaftspolitik: Gedenkenschrift fur Erich Preiser*, Passau: Passavia Universitatsverlag, 1981, pp. 475–508.

______, "Facts and Factors in the Recent Evolution of Business Cycles in the United States," National Bureau of Economic Research (Cambridge, MA) Working Paper No. 2865, 1989.

______ **and Moore, Geoffrey H.,** "The Recession and Recovery of 1973–1976," *Explorations in Economic Research*, Fall 1977, *4*, 471–577.

______ **and** ______, "The Timing and Severity of the 1980 Recession," in G. H. Moore, ed., *Business Cycles, Inflation and Forecasting*, Cambridge, MA: Ballinger, 1983, pp. 11–17.