The results invite direct comparison with those of Kleidon (1986a). As noted in the Introduction, Kleidon argued that the original findings of excess asset-price volatility of Shiller (1979, 1981) and LeRoy and Porter (1981) reflected nothing more than bias induced by faulty trend correction. He suggested that if dividends have a unit root then the original tests should be expected to produce apparent evidence of excess volatility even if the present-value model is true. To demonstrate this he reported simulations assuming that dividends follow a random walk.

Here we accept the random-walk model and specify volatility tests that are valid under this specification. As noted, we find statistically significant excess volatility. Others (Campbell and Shiller [1988] in particular) have reported similar results. However, Campbell and Shiller relied on a log-linearization which, contrary to their evidence, has the capacity to introduce significant approximation error into volatility tests. The fact that we report results similar to Campbell and Shiller (1988) using exact methods rather than approximations adds to the credibility of their results.

*In the May 1990 version of this paper favorable from the authors upon request, we used the Campbell and Shiller (1988) log-linearization to derive an approximate expression for $V(\rho^2|d)$ in the geometric-random-walk case:

$$V(\rho^2|d) = \frac{\beta^2 \sigma^2}{(1 - \beta \mu^2)(1 - \beta \mu)}.$$  

The exact expression (7) derived above, is

$$V(\rho^2|d) = \frac{\beta^2 \sigma^2}{1 - \psi' \left( \frac{\sigma^2 + \sigma^2'}{2} \right) (1 - \beta \mu)}.$$  

The two expressions differ in that $\sigma^2$ does not appear in the denominator under the log-linearization. Substituting estimated parameter values, we find that the log-linearization introduces a 15 percent downward bias in the point estimate of $V(\rho^2|d)$.

REFERENCES


Have Postwar Economic Fluctuations Been Stabilized?

By Francis X. Diebold and Glenn D. Ruderguson*

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. This is no parallel for such a sequence of mild—such as 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 [GDP], 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.

Romer's reinterpretation 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 uncontroversial 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...
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. 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 examining 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 y 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 changing nature of business fluctuations.

In our analysis, however, we do not accept the NBER chronology unquestioningly. One clear truth about 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 is, that the distributions of prewar and postwar durations are identical. We provide empirical results in Section III and offer 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 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). 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.
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; Lowman, 1964; Newell, 1964; George W. Coos, 1963a, b; Zarnowitz, 1963a, b). The differences between the proposed alternatives and the official postwar chronology are minor; of the eight dates examined by Coos, for example, his suggested changes would shift one peak back by one month, another forward by two months, and one trough one month 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 Des L. Stock and Mark W. Watson (1989). They have attempted to formalize the notion that the business cycle is defined by the count 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 data are of varying quality. The dates in the prewar 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 the replacements of one peak 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.

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 evaluations of the early NBER dates, notably Rengdals 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 conclusions 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 represented 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 an annual 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, 1967; 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 conclusions 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
II. A Test of Duration Stabilization

Consider the two samples of prewar and postwar durations of size \( n_1 \) and \( n_2 \), \( \{X_1, \ldots, X_n\} \) and \( \{Y_1, \ldots, Y_n\} \). 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 \( k \)]. These 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 \( \{X_1, \ldots, X_n, Y_1, \ldots, Y_n\} \) by their ranks \( \{R_1, \ldots, R_n\} \), where \( n = n_1 + n_2 \). Then the Wilcoxon test statistic is formed as the sum of the ranks in the second sample:

\[
W = \sum_{i = n_1 + 1}^{n} R_i.
\]

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 the observations) is simply given by

\[
P(R_1 = r_1, \ldots, R_n = r_n) = 1/n!
\]

in the case of a tie, the relevant ranks are replaced by the average of the ranks of the tied observations for all permutations \( (r_1, \ldots, r_n) \) of \( (1, \ldots, n) \).

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).

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,

\[
t = \frac{(n_1 - n_2)/n}{\sqrt{\frac{\sum_{i = n_1 + 1}^{n} (X_i - \bar{X})^2}{n_1 - 1} + \frac{\sum_{j = 1}^{n_2} (Y_j - \bar{Y})^2}{n_2 - 1}}}
\]

where

\[
s^2 = \frac{(n_1 - 2)/n}{(n_1 - 1)} \times \left( \frac{\sum_{i = n_1 + 1}^{n} (R_i - \bar{R})^2}{n_1 - 1} \right) \cdot \left( \frac{\sum_{j = 1}^{n_2} (R_j - \bar{R})^2}{n_2 - 1} \right)^{1/2}
\]

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.

Under the maintained assumption that the distributions of durations differ only by a shift in location \([i.e., G(k) = F(k + \Delta) \text{ for all } k]\), we can also produce a confidence interval for the location shift, \( \Delta \). Consider the \( n_1 \) element sequence of differences \( \{D_{1i}, i = 1, \ldots, n_1\} \), and order them so that \( D_{1i} < D_{12} < \ldots < D_{1n_1} \). For a given significance level \( \alpha \), let \( k_\alpha \) be an integer defined from the confidence interval

\[
P(k_\alpha < U \leq n_2, n_2 - k_\alpha) = 1 - \alpha
\]

where

\[
U = W - n_1(n_1 + 1)/2
\]

Critical values are also tabulated in John V. Bradley (1986) for \( n_1, n_2 \leq 25 \).

\[
(6) \quad s^* = (n - 2)^{-1/2} \times \left( \frac{\sum_{i = n_1 + 1}^{n} (R_i - \bar{R})^2}{n_1 - 1} \right) \cdot \left( \frac{\sum_{j = 1}^{n_2} (R_j - \bar{R})^2}{n_2 - 1} \right)^{1/2}
\]

\[\text{Vol. 82 No. 4} \quad \text{Diebold and Rudebusch: Postwar Economic Fluctuations}\]

\[\text{The American Economic Review} \quad \text{September 1992}\]

\[\text{998}\]

\[\text{Table 2—Listing of Duration Samples}\]

\[\begin{array}{l}
\text{A. Pre-World War II (December 1954—June 1980)}
\end{array}\]

\[\begin{array}{l}
\text{A1: All observations} \\
\text{A2: Excluding observations before May 1955} \\
\text{A3: A1, eliminating 1987 and 1989 contractions} \\
\text{A4: A1, excluding wartime observations} \\
\text{A5: A2, excluding wartime observations} \\
\text{A6: A3, excluding wartime observations} \\
\end{array}\]

\[\begin{array}{l}
\text{B. Pre-Great Depression (December 1954—August 1929)}
\end{array}\]

\[\begin{array}{l}
\text{B1: All observations} \\
\text{B2: Excluding observations before May 1955} \\
\text{B3: B1, eliminating 1987 and 1989 contractions} \\
\text{B4: B1, excluding wartime observations} \\
\text{B5: B2, excluding wartime observations} \\
\text{B6: B3, excluding wartime observations} \\
\end{array}\]

\[\begin{array}{l}
\text{C. Pre-World War I (December 1954—December 1914)}
\end{array}\]

\[\begin{array}{l}
\text{C1: All observations} \\
\text{C2: Excluding observations before May 1955} \\
\text{C3: C2, eliminating 1987 and 1989 contractions} \\
\text{C4: C2, excluding wartime observations} \\
\text{C5: C3, excluding wartime observations} \\
\end{array}\]

\[\begin{array}{l}
\text{D. Post-World War II (February 1945—July 1990)}
\end{array}\]

\[\begin{array}{l}
\text{Z: All observations} \\
\text{Z*: Excluding wartime observations} \\
\end{array}\]

Turning-point dates in order to avoid potentially unreliable data in the very early period and Gil elimination of the 1887 and 1899 recessions 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.)

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

8See Peter J. Bickel and Ira 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 $W$. Then it can be shown (Bickel and Doksum, 1977) that

$$
(9) \quad P\left(D_{i_{2}, v_{2}} \leq \Delta \leq D_{i_{2}, v_{1}} \right) = 1 - \alpha.
$$

Thus, a two-sided $(1 - \alpha)$-percent confidence interval for $\Delta$ is $(D_{i_{2}, v_{2}}, D_{i_{2}, v_{1}})$. Alternatively, the two $(1 - \alpha)$-percent one-sided confidence intervals are $(D_{i_{2}, v_{2}}, \infty)$ and $(-\infty, D_{i_{2}, v_{1}})$. 

### 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

Table 3 - Mean Growth Rate of Real GNP

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean duration (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postwar sample: 1940-1990 (Z)</td>
<td>0.025</td>
</tr>
<tr>
<td>Pre-war sample: 1910-1938 (A1)</td>
<td>0.011</td>
</tr>
<tr>
<td>1939-1945 (A2)</td>
<td>0.012</td>
</tr>
<tr>
<td>1946-1990 (Z)</td>
<td>0.012</td>
</tr>
<tr>
<td>1910-1914 (1)</td>
<td>0.013</td>
</tr>
<tr>
<td>1915-1919 (2)</td>
<td>0.013</td>
</tr>
<tr>
<td>1920-1924 (3)</td>
<td>0.014</td>
</tr>
<tr>
<td>1925-1929 (4)</td>
<td>0.015</td>
</tr>
<tr>
<td>1930-1934 (5)</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Note: The real GNP sample from 1909 to 1929 comes from Romer (1984, pp. 22-3); the data after from the national income and product accounts (NIPA).

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. Thus, any evidence for duration stabilization does not reflect changes in trend growth.

With the 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. For example, the top row of Table 4 compares the prewar expansion sample $A_1$ (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 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 test rejects the null hypothesis of no stabilization in favor of longer postwar expansions or shorter postwar contractions.
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 of its prewar counterpart.

The results are very different for whole cycles, whether measured from trough to peak or peak to peak. Table 6 provides the statistics for cycles measured from peak to peak; similar results were obtained for trough-to-peak 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 criticism 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 stabilizing policy hypotheses. 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, a consistent 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 durations grow without bound, and expected contraction durations collapse to zero. 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 postshift 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 to lead to duration stabilization.

REFERENCES


... and Tanizaki, Hisashi, "On the Comparative Size and Power Proper-

*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.


