Volume Title: Forecasting and Recognizing Business Cycle Turning Points

Volume Author/Editor: Rendigs Fels and C. Elton Hinshaw

Volume Publisher: NBER

Volume ISBN: 0-870-14479-0

Volume URL: http://www.nber.org/books/fels68-1

Publication Date: 1968

Chapter Title: 1948–61: Accuracy of Dating

Chapter Author: Rendigs Fels

Chapter URL: http://www.nber.org/chapters/c1087

Chapter pages in book: (p. 13 - 18)

conformity with the forecast in question would, on the whole, have been a step in the right direction or in the wrong one." [16] Predictions were assumed to apply to no more than eight months into the future unless the forecaster indicated otherwise. A forecast of conditions in a special sector was judged only by events in that sector. Forecasts of general business were checked either against the composite index of business activity specified by the service or, failing that, against three composite indexes (those of the Annalist, A.T.&T., and the Federal Reserve Bank of New York).

The relation of Cox's scores to recognition of cyclical turning points is not clear. The scores during the six months before and after a cyclical peak or trough may or may not pertain to forecasts of the turning points in question. Nevertheless, Chart I-3 is presented to supplement Chart I-2. It displays some tendency toward increasing correctness of forecasts as turns are approached and passed, but the tendency is less marked for the averages than in Chart I-2. For 1929, the scores fluctuate from month to month in a highly erratic fashion.

4

1948–61: Accuracy of Dating

For each of the eight turning points between 1948 and 1961, we have studied reports published by a number of contemporary observers. Like Cox, we first excerpted short quotations from current forecasts. We scored the excerpts in two different ways, for accuracy of dating and for degree of recognition. The scores for dating range from 0 to 100. The maximum score was given for designating a peak or trough within one month of the NBER date. A forecast that missed by two months received a score of 75; by three months, 50; by four months, 25. Thus, positive scores were given for forecasts of a peak or trough anywhere within a nine-month interval centered on the NBER reference date. A

[16] *Ibid.*, p. 19.

score of zero means either a forecast that missed by more than four months or, much more frequently, no forecast of dating. Errors in dating a turn too early were scored the same as dating them too late. From the point of view of forecasting, and also for policy-making, the latter are more serious.

As just described, the scoring system for dating might appear to be objective, but in practice it involved much subjective judgment. Most of the analysts studied gave little explicit attention to determining the date of the peak or trough.[17] As a result the scores were often based on a chance statement that might or might not be inferred to forecast the date of the turn. Close questions sometimes arose as to whether a forecast should be given 0 or some positive score (often 100), 0 being given if the analyst was deemed to have made no forecast of dating, a positive score if an oblique statement could be inferred to have forecast a turn within the nine-month target period. If a publication predicted the date of a peak or trough in one issue but neither revised nor repeated it in the next, we assumed that the forecast of dating remained the same and gave the new forecast the same score as the old. This procedure meant that a close decision on a chance statement sometimes made a large difference to the total score of a given analyst.[18]

[17] Evidence of lack of interest in dating is afforded by the high proportion of scores based on no forecast at all—63 per cent. (As noted above, a publication making no forecast normally received a score of zero, but, as noted below, a publication that received a positive score one month but said nothing thereafter was given the same positive score every month until six months after the peak or trough.)

[18] The analysts were first scored by the author and C. Elton Hinshaw independently. Where differences could not be resolved by discussion, the final decision was made by the author. Predictions about timing seldom designated a specific month and sometimes were quite vague. Such indefiniteness was penalized. The penalty consisted of a deduction of 12½ for each month by which the period in which the turn was predicted to occur exceeded three months. Under this system, a forecast of a peak in the first half of 1960 would receive a score of 50. The NBER reference date is May. The midpoint of the first half-year is March 31–April 1, just outside the target period of April-May-June, and would be scored 87½ minus a penalty of 37½ for indefiniteness. A forecast of a peak in the three-month period March-April-May, on the other hand, would get a score of 100. The midpoint of the forecast period falls within the three-month target period, and no penalty would be assessed for indefiniteness. If the forecast was vague about when the turn was expected, the predicted timing was inferred as well as possible, and then a penalty was assessed for the vagueness. The decision to adopt an explicit scoring system was made after the quotations had been excerpted from the publications. A spot check has shown that the quotations were not always the best possible for purposes of scoring. We did not have excerpts from every issue of the weekly publications. The missing weeks were treated in

Eight of the publications studied are available monthly (or weekly) for all eight turns between 1948 and 1961. Averages for these eight constitute the core of our study. They include four business publications, two security services, and two bank letters. In terms of circulation and influence, they constitute an elite group. Three of the business publications are the best known names in the field. The two leading security services are represented, and the bank letters come from large, highly respected institutions. Two publications that relied principally on the business cycle indicators developed by the NBER were also scored.[19] Since their records did not cover the entire eight turns, averages including them are shown separately.

Panel A of Chart I-4 summarizes the results. The monthly averages for eight analysts for four troughs and (to a lesser extent) for four peaks show the expected upward drift. Three months before the turn the averages are very low (so low that there seemed little point in scoring earlier months). The scores for accuracy of dating are much higher at troughs than at peaks. The average of the 320 scores at the four peaks (eight forecasters, ten months per peak) is only 10 compared with 37½ for troughs. The average for the sixth month after peaks is only 23 compared with 71 for the sixth month after troughs. Of 360 scores associated with peaks for all ten publications, 303 are zero (84 per cent). At troughs, there are only 142 zero scores out of 370 (38 per cent). That is, there were four times as many nonzero scores at troughs as at peaks (228 vs. 57).

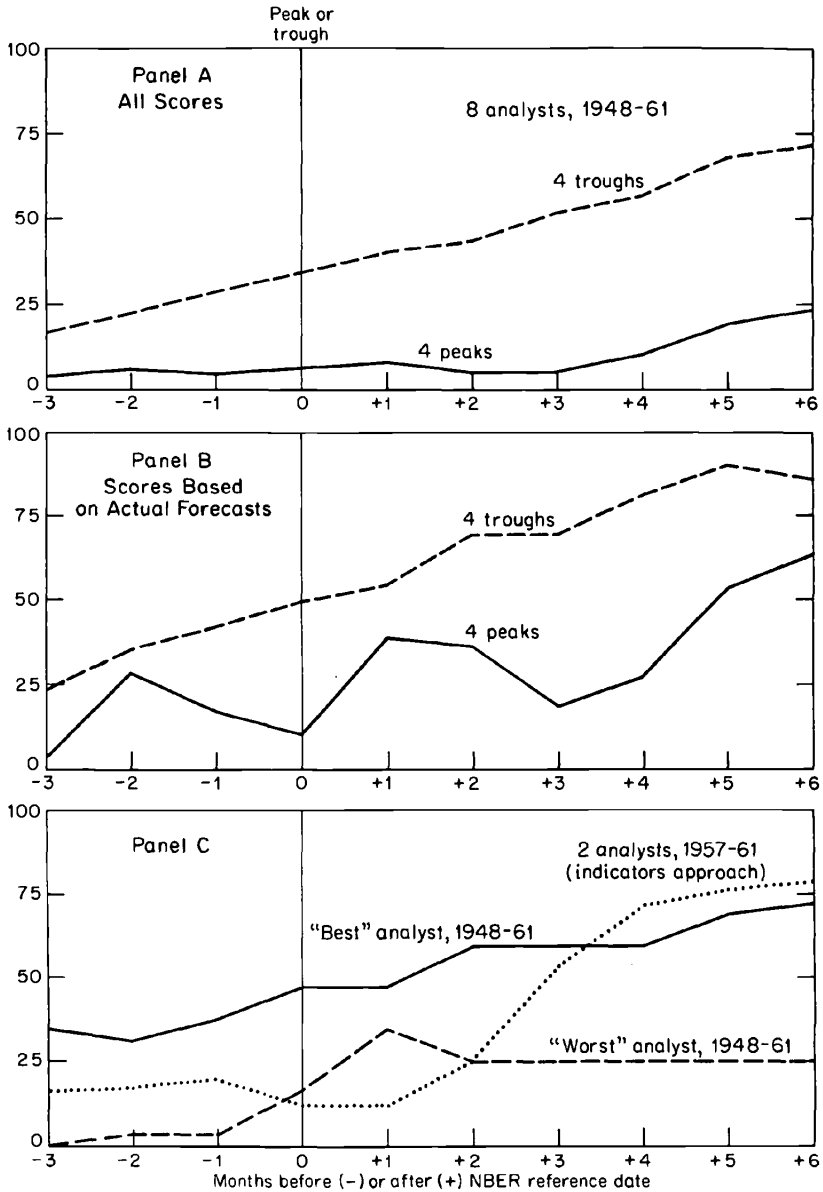
The scores in Panel A of Chart I-4 probably understate the ability of the eight publications to forecast dating. For all ten publications, only 37 per cent of the scores are based on actual forecasts. In 50 per cent of the cases, a zero score was assigned because no forecast of the date

the same manner as cases where the publication said nothing about timing, that is, the previous score was carried forward. The weekly scores were then averaged to get the monthly scores. The forecasts of one monthly publication were regarded as having been made in the month preceding the date of publication, since it regularly appears toward the end of the preceding month. Otherwise no effort was made to allow for the fact that the forecasts of different publications were made at different times during the month.

[19] It is the policy of the National Bureau of Economic Research not to identify the sources of forecasts evaluated in the large research project of which this study is a part. The policy applies to forecasts that have been published as well as unpublished forecasts given to the NBER on a confidential basis. The reasons for this policy are given by Zarnowitz in *An Appraisal of Short-Term Economic Forecasts*, pp. 1–2.

CHART I-4

Accuracy of Dating Cyclical Peaks and Troughs, Ten Analysts, 1948–61



SOURCE: Appendix I, Table B.

of the turn had been made. (In 10 per cent of the cases, a score of zero was given because the forecast of the date was poor.) The remaining 13 per cent of the scores are positive scores carried forward from an earlier month. These figures, as indicated, include the two publications not available for all turns. These two made forecasts of dating more frequently than the others. For the eight alone, only 33 per cent of the scores are based on actual forecasts.

Panel B of Chart I-4 shows the results of excluding from the averages all of the zero scores based on no forecasts and a majority of the positive scores based on no forecasts. In a few cases, a positive score was retained because the publication clearly had no reason to repeat a conclusion already reached. Many of these cases occurred in 1961, when most forecasters knew within a few months that the trough had occurred in February, lost interest in the matter, and went on to discuss how vigorous the upswing would be. Positive scores not based on an actual forecast were retained only if three conditions were met: (1) the score was for a date no earlier than the second month after the NBER peak or trough; (2) the date given was not subsequently revised by the publication during the scoring period (i.e., within six months of the turn); (3) the accompanying certainty score (see the next chapter) was at least 75. If these three conditions were not met, failure to repeat an earlier forecast might have resulted from a change of opinion.

Since most of the scores thus excluded were zeros, the patterns in Panel B are higher than those in Panel A. The scores continue to be higher at troughs than at peaks. Panel B presumably overstates the ability of the eight publications to forecast the dates of turns, since failure to forecast a date can result from inability to do so. A fair assessment of forecasting ability probably lies somewhere between the patterns of the two panels. The data going into Panel A, however, have the advantage of greater comparability: for every month of every scoring period there are eight observations. The data for Panel B include a number of observations for each month varying from zero to eight.[20]

[20] The variation in number of observations means that a choice had to be made between two weighting systems in calculating the averages shown in Chart I-4, Panel B (the data for these can be found in Appendix I, Table B). Giving equal weight to each published forecast gives an upward bias, since more analysts give forecasts of dating in easy cases like 1961 (when only 8 per cent of the scores were excluded under our rules) than at turns hard to forecast like 1948 (when 89 per cent of the scores had to be dropped). The alternative of weighting each month equally regardless of the number of scores entering into that month's

For the sake of comparability, the dating scores discussed in the remainder of this paper include all scores whether based on forecasts or not.

There was a wide variation among the scores for individual analysts. For each analyst all the scores for the full ten-month period and for all eight turns were averaged. The averages ranged from 16 to 52. The median of the average scores of the eight analysts was 23, the mean 26. Panel C of Chart I-4 contrasts the patterns of the "best" and "worst" analysts. It also shows the average pattern for 1957–61 for the two publications that relied heavily on business cycle indicators.

A word of caution is in order about the "best" and the "worst" forecaster in Chart I-4, in other charts in my study, and in Hinshaw's comparison of the "best" with the Federal Open Market Committee. The "worst" analyst is the one with the lowest average score for degree of certainty (see the next chapter) but is only the second lowest with respect to accuracy of dating. The "best" analyst has the highest average score for both accuracy of dating and degree of certainty, taking peaks and troughs together. It has the highest scores, moreover, for peaks alone with respect to dating and for troughs alone with respect to both dating and certainty, but not for peaks alone with respect to certainty. In fact, its certainty scores at peaks were erratic—it had the highest score at one peak, the fourth highest at another peak, and the seventh highest (i.e., second worst) at the other two.[21] As a result, the so-called "best" displayed the greatest variability in certainty scores of the eight forecasters, even when all peaks and troughs are taken together. Since consistency is a virtue in forecasting, these considerations throw serious doubt as to whether the "best" can be considered a superior forecaster. Even its superiority with respect to dating is open to question on grounds that its higher scores resulted from the others ignoring the question of determining the date of the peak or trough. Of course, the range (or the uppermost observation) is in any event of limited value, since the range depends on the size of the sample. The best in a sample of eight is likely to be inferior to the best in a sample of eighty but superior to the best in a sample of four.

average (unless the number was zero) was therefore used instead. (A chart comparing the two procedures shows that the results are closely similar at troughs. At peaks the expected bias is evident but not great.)

[21] See Appendix I, Table J.