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: Degree of Certainty

Chapter Author: Rendigs Fels

Chapter URL: http://www.nber.org/chapters/c1088

Chapter pages in book: (p. 19 - 24)

5

1948–61: Degree of Certainty

The quotations excerpted from the publications of the various analysts were scored for degree of certainty as well as for accuracy of dating. The user of forecasts needs to know not only what the forecast is but also the degree of confidence the forecaster has in it. Unfortunately, commentators often take refuge in ambiguity instead of giving odds on various precisely defined possibilities.

A general practice of communicating the degree of confidence in a forecast by stating subjective odds would not only be a clearer way of describing the outlook to the reader but would also facilitate retrospective appraisal of the record of predictions. For the specific purposes of the present study, increasing recognition of an imminent or recent peak would be revealed by rising odds on a downturn, and increasing recognition of a trough by rising odds on an upturn.[22]

The scoring system chosen was based on the odds on a cyclical turn implicit in the language of the forecast. Each quotation was scored for degree of recognition on a scale from 0 to 100 (since we only used multiples of five, there were twenty-one possible scores). A score of 50 means a roughly even chance of a cyclical turning point. A score of 100 represents virtual certainty (strictly speaking, chances greater than 97½ per cent).[23] As indicated earlier, each quotation was scored indepen-

[22] If all forecasters gave odds, in time a large enough sample would be generated to make possible a test of the hypothesis that half of the forecasts stated to have five chances in ten come true, 60 per cent of those given six chances in ten, etc. Such data, however, are too rare to make such a test.

[23] No corresponding statement can be made about low scores, which are biased downward compared to the implicit odds on a turning point. In the absence of any evidence to the contrary, we assumed that a publication regarded the odds of a turn in the target period as "normal." The target period is a three-month interval centered on the NBER reference date. Since recent expansions have averaged between two and three years in length (somewhat less in earlier times), the "normal" odds of a peak in any three-month interval during a period known to be an expansion may be taken as one in ten or a little less. Accordingly, we gave a publication a score of 10 in the vicinity of peaks unless it implied that

dently by C. Elton Hinshaw and myself, and discrepancies were resolved by discussion as much as possible. When a doubt remained, the two original scores were averaged.[24] The resulting scores are subjective. Although in a majority of cases the two initial (independent) scores were close together, in a disconcerting number of cases they differed widely. The same person scoring the same series of forecasts after a lapse of several months sometimes gave quite different scores.

As a result of both the nature of the materials and the procedures followed, the scores are more reliable for tracing the pattern of increasing recognition by a single forecaster than for comparing the forecasts of different publications. Quite subtle shifts in wording and emphasis from one month to the next by a given publication often revealed fine shades of increasing uncertainty about prospects for continued expansion in the vicinity of peaks (and similarly for troughs); this is why as many as twenty-one different possible scores were needed. Even so, the scores could not always reflect discernible changes in the attitudes of the forecasters. On the other hand, the difference in language of different publications was often hard to interpret. It was much easier to tell whether a given forecaster had become more or less optimistic since last month than to tell whether he was more or less optimistic than his fellows. Such difficulties, however, are minimal in the vicinity of 50 and 100. If a forecaster expresses great certainty, the only question is whether the score should be 95 or 100. Similarly, it is comparatively easy to tell whether a forecaster thinks a turn more (or less) likely than not. Therefore, errors in scoring on the wrong side of 50 are unlikely. This is fortunate because 50 can be regarded as the dividing line between recognition and failure to recognize.

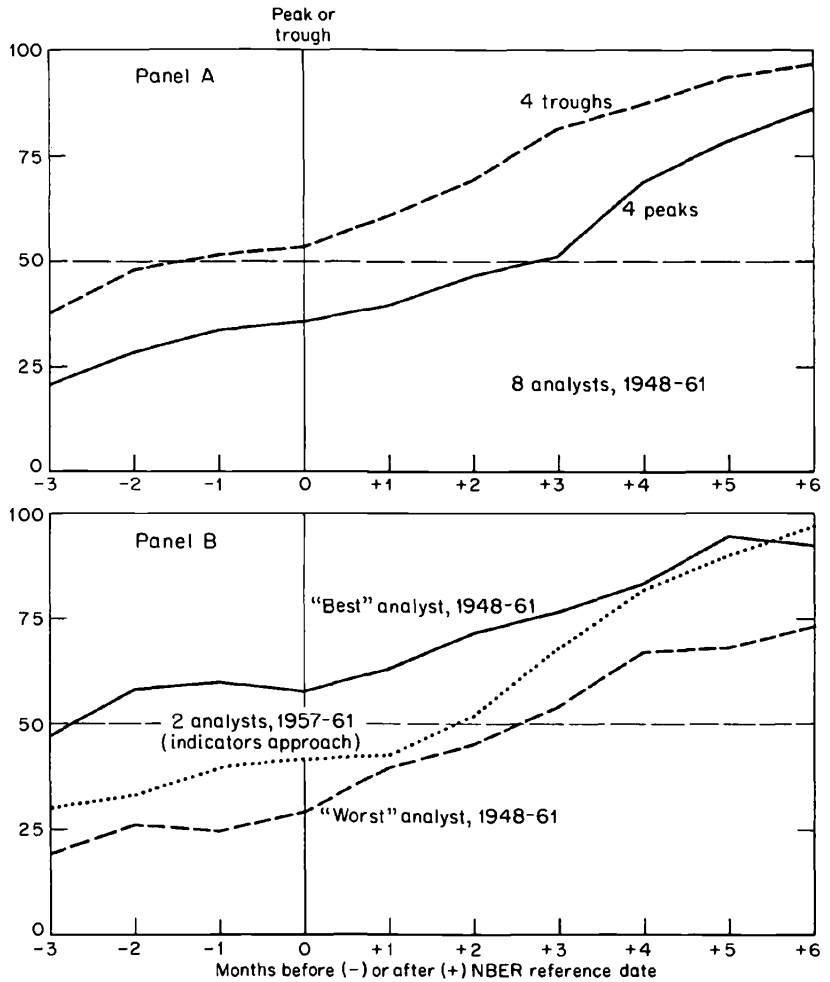
Our procedure was to score the ten (or more) quotations for a given forecaster in succession before scoring other forecasters for the same peak or trough. This procedure accentuated the inherent tendency for the pattern of scores for a given forecaster to be more reliable than comparisons among different forecasters. Despite the weaknesses of the scores, they are reliable enough for the purposes to which they are put

the odds were greater or less than "normal." Since contractions average shorter than expansions, in the case of troughs we gave a score of 15 unless the publication implied that the odds were greater or less than normal. Defining the target period as three months, however, turned out to be too restrictive and was not adhered to rigidly for scores of 50 and higher.

[24] In the case of weekly publications, the weekly scores for each month were averaged. Where we did not have quotations for a given week, we interpolated.

CHART I-5

Degree of Certainty of Forecasts of Cyclical Peaks and Troughs, Ten Analysts, 1948–61



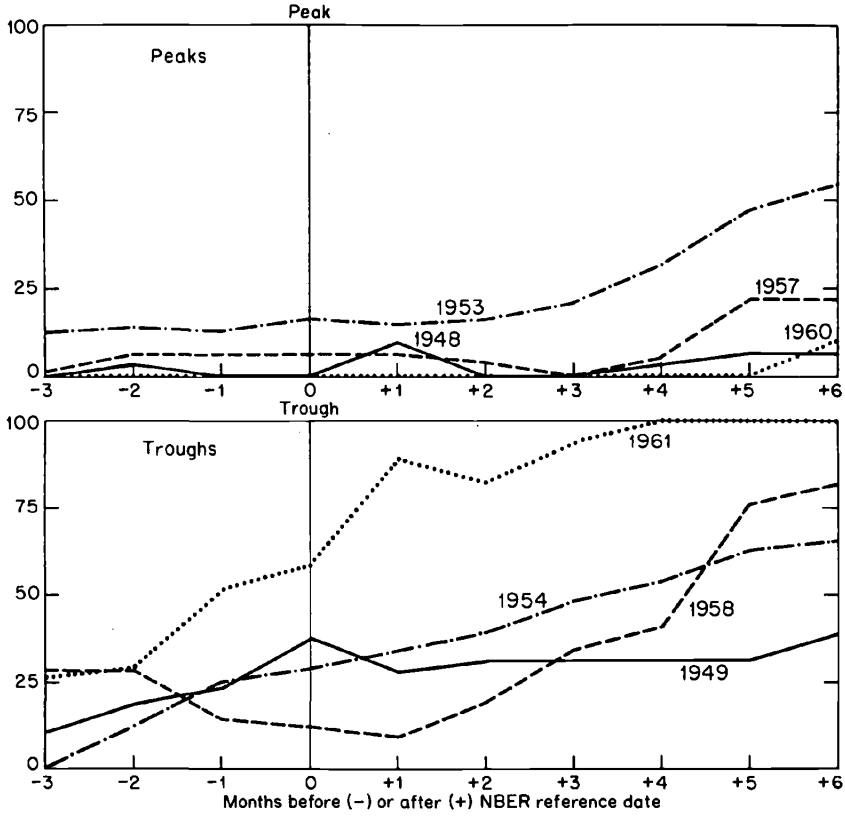
SOURCE: Appendix I, Table C.

here. But the reader should be cautioned against using the findings for more ambitious purposes.

Chart I-5 exhibits the expected pattern of increasing certainty over time. Panel A shows a marked difference between peaks and troughs.

CHART I-6

Comparison of Scores for Accuracy of Dating at Four Peaks and Four Troughs, 1948–61
(averages of eight analysts)

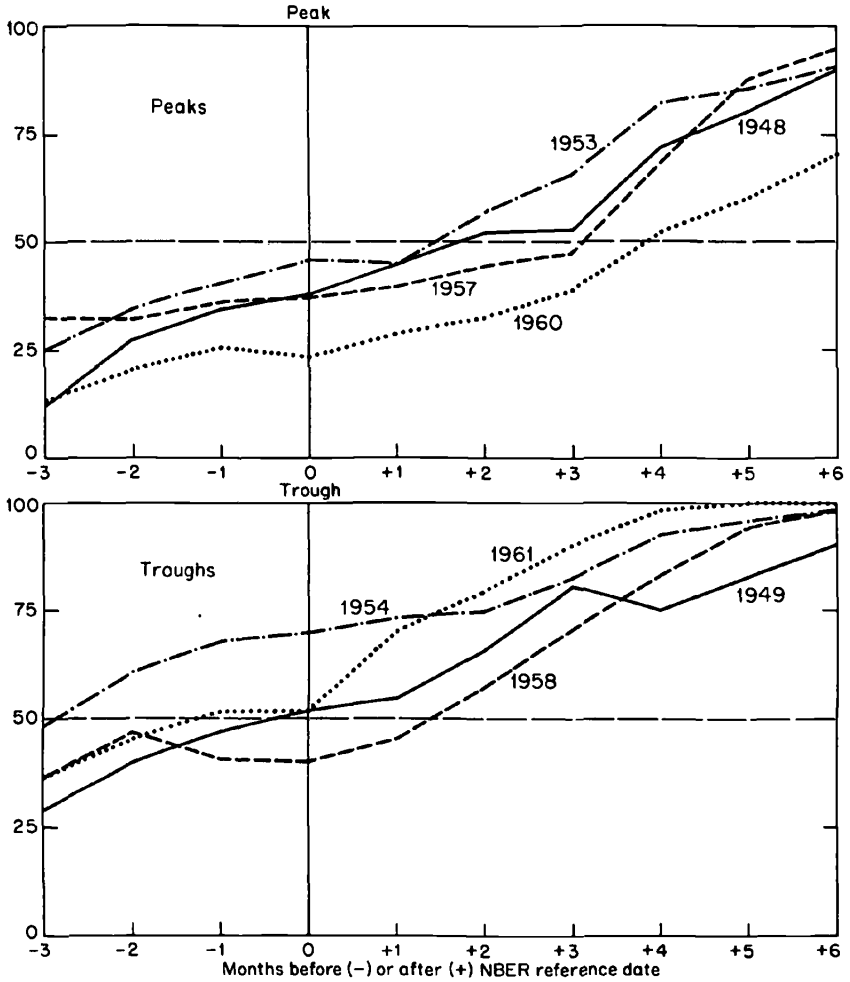


SOURCE: Appendix I, Table D.

At troughs, the mean score for eight forecasters first exceeds 50 (fifty-fifty odds) one month before the NBER reference date. At peaks, a mean of 50 is not achieved until three months after the reference dates. If confirmation is defined to mean an average score of 95 (95 per cent confidence), the group of analysts studied were able to confirm that a trough had occurred six months after the event but could not do so well for peaks. Averaging peaks and troughs together, only one of the ten analysts had an average score greater than 50 two months before the

CHART I-7

Comparison of Scores for Degree of Certainty at Four Peaks and Four Troughs, 1948–61
(averages of eight analysts)



SOURCE: Appendix I, Table E.

reference date. Four out of eleven achieved an average score of 95 five months after the reference date (counting peaks and troughs together),[25] none earlier. If all scores for all months are averaged for each forecaster, the scores range from 44 to 71. The same publication that scored highest for timing also scored highest for degree of certainty. Panel B of Chart I-5 compares the "best" and the "worst" analysts for 1948–61. It also shows the average pattern for 1957–61 of the two publications that depended heavily on business cycle indicators.

All the average scores, for both accuracy of dating and degree of certainty, are subject to a serious limitation. As Charts I-6 and I-7 show, there is great variation among the scores for different turning points, not only the systematic differences between peaks and troughs already noted, but also differences among individual peaks and individual troughs. Discussion of each turn is in order.

6

The Recognition Pattern: A Chronological Review

The 1948 Peak

Scores for both dating and certainty run lower for peaks than for troughs, and among the four peaks since World War II, they run lower for 1948 than for any other year except 1960. Although there is nothing to choose between the certainty patterns for 1948 and 1957 shown in Chart I-7, the 1948 scores for dating are noticeably lower (see Chart I-6). Of the 80 scores for dating near the 1948 peak (eight forecasters, ten months), 75 were zero. (Seventy-one were zero because no forecast was made.)

[25] This includes one publication not used in the averages shown in Charts I-4 and I-5. The omission of this publication reduces the number with an average of 95 five months after the reference date to three in ten.